

# THIS WEEK

## EDITORIALS

**CAREERS** Nature launches mentoring awards in Ireland **p.8**

**WORLD VIEW** Discovery of gravitational waves put on hold **p.9**

**CLIMATE** Coral reefs help keep damselfish from distress **p.10**



## Don't feed the trolls

*Patent abuse slows down research and innovation, and must be confronted. Delays to US legislation are not reassuring, but there has been some progress in the courts.*

Not long after news of his experiments got out, the trolls came for Rob Carlson. A consultant in Seattle, Washington, Carlson is part of a growing movement of biohackers who tinker with biotechnology in their garages. But when word reached the media several years ago that Carlson intended to commercialize his inventions, he was threatened with lawsuits from people who claimed to hold patents that covered the entire field of proteomics — the study of the proteins that make up cells and organisms. If Carlson did not pay a substantial settlement, the purported patent holders said, they would sue him for infringement.

And so Carlson was introduced to the world of patent trolls, a pejorative term for people or organizations who file or license patents solely to use them to extort money from firms that infringe them. In 2012, patent trolls accounted for 62% of all patent legislation in the United States; in 2011, such cases cost companies US\$29 billion. Attempts to rein in patent abusers are mounting. On 2 June, the Supreme Court issued a decision that tightens requirements for patent claims to be clear and unambiguous, potentially limiting the broad claims that foster abuse. US President Barack Obama has made tackling trolls a priority, and last December, the House of Representatives passed a bill proposing an 'Innovation Act', which included measures to counteract the problem.

But on 21 May, Senator Patrick Leahy (Democrat, Vermont), chair of the Senate Judiciary Committee, announced that he was taking the bill off the committee's agenda. He said that nearly a year of hard work had failed to produce legislation that would temper trolls without harming genuine patent holders. With Congress heading into an election in November, it is unlikely that the bill will be resurrected this year.

The legislation's demise highlights how hard it is to design patent policy that satisfies two large technological domains. Trolls mainly target technology firms, in part because patents on software and business methods are often broad. Many technology companies already find patents to be a nuisance — particularly those that make broad claims on business methods or software (see *Nature* **509**, 152–154; 2014).

By contrast, the biotechnology and pharmaceutical industries hold their patents dear: intellectual-property protection can be important during the often lengthy struggle to win regulatory approval for a drug or genetically engineered crop. Trolls have not so far tended to trouble these industries, but that may change. In a study released this year, researchers found dozens of university-held patents that could be deployed against bioscience companies, including some that cover methods to screen for or manufacture new drugs (R. Feldman and W. N. Price *UC Hastings Research Paper No. 93* <http://doi.org/s2m>; 2014).

Universities also value patents, both to encourage the commercialization of academic inventions and as a source of revenue. Academic technology-transfer offices tend to raise the bulk of their funds from licensing biomedical patents. In April, the Association of University Technology Managers joined groups including the Biotechnology Industry Organization in signing a letter to Leahy, opposing

the proposed Innovation Act. The bill would make it difficult and expensive to enforce their patents, they said.

University opposition to the Innovation Act has fuelled claims that some academic institutions have themselves become patent trolls. Industry insiders have made this assertion many times over the years, particularly as universities have become more aggressive in protecting their patent holdings by suing potential infringers. A popular term for a patent troll is a 'non-practising entity': a party that does not intend to market products based on its patents. By this very broad definition, universities — which license their patents instead of marketing products directly — would be patent abusers. But of course they are not, as long as they hold their mission to help society above their drive to bring in cash. Academic institutions need to make their priorities clear: the practice of licensing patents to trolls to raise funds does not help their public image (see *Nature* **501**, 471–472; 2013).

It is disappointing that Congress will do nothing in the near future to slow the steady march of the patent troll. But luckily, legislation is not the only option. By the end of June, the Supreme Court is expected to rule in a complex patent case that could narrow the scope of software and business-methods patents (see *Nature* **507**, 410–411; 2014). The US Patent and Trademark Office has initiatives to make it easier to determine who owns a patent. And the US Federal Trade Commission is studying troll behaviour. If the target is better defined, it may well become easier to design legislation that hits the mark. ■

**“Nearly a year of hard work failed to produce legislation that would temper trolls without harming genuine patent holders.”**

## Renewed energy

*Reforms at the US Department of Energy are recharging research.*

When physicist Steven Chu took over as head of the US Department of Energy (DOE) in 2009, he vowed to reform its research culture. Many felt that the department had become much too bureaucratic — too rigid, too unresponsive to new opportunities, too divided into disciplines and too isolated from the needs of the marketplace.

The following year, Chu launched five Energy Innovation Hubs intended to mimic the research style that he remembered from his time working at the AT&T Bell Labs in Murray Hill, New Jersey. Each hub would focus on a well-defined challenge in the area of renewable energy

— a top priority for the then-new administration of US President Barack Obama. It would bring together all the necessary expertise, from basic and applied research to engineering and early product development.

Four years later, there is justified, if cautious, optimism about the outcome of Chu's experiment. Viewed purely as research projects, most of the hubs seem to be doing well. In the next few months, the Joint Center for Artificial Photosynthesis, headquartered at the California Institute of Technology in Pasadena, hopes to demonstrate a first-generation prototype of an 'artificial leaf' — a cheap, robust and highly efficient system able to make liquid fuels out of sunlight, air and water (see page 22). The Joint Center for Energy Storage Research, headquartered at the DOE's Argonne National Laboratory near Chicago in Illinois, is likewise making good progress towards its goal: devices that can store much more electricity in much less space than the current champions, lithium-ion batteries (see *Nature* 507, 26–28; 2014).

Only one of the five hubs has fallen by the wayside. The Energy Efficient Buildings hub, headquartered in Philadelphia, Pennsylvania, was eventually judged to be too diffuse in its goals for DOE purposes, and too oriented towards trying to get people to use currently available technology. But it still exists. In April it took a new name — the Consortium for Building Energy Innovation — and relaunched itself as an independent research and demonstration centre.

There are also grounds for optimism about the hubs' larger purpose of transforming the DOE research culture — although in this case, the progress is less clear-cut. In some ways the agency is as bureaucratic as ever. And talk of change within the department has provoked its share of resistance from individuals who feel that their programmes are threatened.

Nevertheless, there is considerable excitement in the DOE — a sense of new opportunities, new ventures, new people. The hubs are responsible for some of that feeling, as are innovations such as the Advanced Research Projects Agency — Energy (ARPA-E), established in 2009 to fund speculative, high-risk, high-reward investigations, and a network of Energy Frontier Research Centers, launched the same year to promote cutting-edge basic research.

But at least as important is the sense that the people at the top understand and support reform. Chu's initiatives have been continued by

his successor, physicist Ernest Moniz — who last year told Congress that the hubs would be a good model for reforming the DOE's network of 17 national laboratories. Last month, Moniz appointed a panel to review the national labs, with a report due early next year.

Obama's administration has been supportive. In both his 2013 and 2014 State of the Union addresses, Obama called for a US\$1-billion National Network for Manufacturing Innovation. An interagency programme modelled in part on the DOE's energy hubs, this would comprise 15 or more centres looking to cut the energy, time and materials required to make things. The goal is to help US industries to compete with low-cost factories in emerging nations such as China, and to make it easier for start-up companies — including many renewable-energy firms — to bring new products to market. Congress has not yet acted on this proposal, but the administration has established several centres using existing funds from the DOE and other agencies.

Such efforts need to be supported and encouraged — especially by Congress, which holds the federal purse strings, and by the energy industry, which can tap vast amounts of cash for activities it perceives to be in its interest. And even here there is reason for optimism. Despite the ideological warfare that has riven Washington DC in recent years, both parties have generally endorsed the DOE's reform efforts. And industry leaders seem ready to work closely with researchers to bring innovative products to market. One example is the Clean Energy Trust, a Chicago-based consortium of energy companies that supports renewable-energy start-ups.

Congress and the Obama administration could greatly help this movement by reviving the idea of the Clean Energy Deployment Administration: a 'green bank' that would pool public and private money for large-scale investments in clean-energy infrastructure. The idea was proposed a few years ago, but abandoned amid budget wrangles. Now that the federal deficit is easing and the economy has begun to improve, it could find renewed support on both sides of the aisle. The future, for once, is starting to look brighter. ■

**“There is considerable excitement in the Department of Energy — a sense of new opportunities, new ventures, new people.”**

## Integrity mentors

*Policies in Ireland and China make Nature's 2014 mentoring awards timely.*

Last month the Chinese Academy of Sciences issued a powerfully worded statement *Towards Excellence in Science* ([go.nature.com/pnhi9k](http://go.nature.com/pnhi9k)). In encouraging a scientific culture of challenging the status quo, it includes a passage that speaks to laboratory behaviour: “To achieve scientific excellence, the scientific community needs to consciously advocate and uphold the scientific spirit, promote the value and focus of science in seeking truth and innovation, establish management structures and mechanisms that suit the characteristics and rules of scientific research, and discourage scientific behaviour aimed only at short-term success or individual benefits.”

This week, the Irish Universities Association has issued a Concordat on research integrity, which includes mention of two aspects (among several) of scientific behaviour needing support: “reliability in performing research (meticulous, careful and attentive to detail), and in communication of the results (fair and full and unbiased reporting), and objectivity: interpretations and conclusions must be founded on facts and data capable of proof and secondary review; there should be transparency in the collection, analysis and

interpretation of data, and verifiability of the scientific reasoning.”

Such statements could all too easily be ignored unless they have teeth. In that spirit, readers might do well to focus on a clause in a document produced by Science Foundation Ireland (SFI), the country's main funding agency and a collaborator on the integrity Concordat. On page 32 of its strategic plan, *Agenda 2020*, is the statement that research integrity will be scrutinized by external audits ([go.nature.com/xjudiz](http://go.nature.com/xjudiz)). Congratulations to the SFI for showing more determination than most to back words with actions.

Excellent science requires, not least, a capacity for researchers to be ruthlessly self-critical — in other words, assuring technical integrity. On discovering something interesting, they need to assume at the outset that they are deluded — that the combination of their object of study and their experimental, or theoretical or simulation set-up is conspiring to make them mistakenly believe that they have a startling new insight to offer an admiring world. They need to show their analyses or data to trusted but critically minded colleagues, in order to avoid mistakes and cherry-picking. Such a culture is best bred by tough but supportive laboratory mentors. In its annual mentoring awards, which has been held since 2005, *Nature* has rewarded outstanding mentors in many countries and regions.

Given Ireland's evident determination to sustain best practices, it is timely that this year's mentoring competition is for scientists resident in that country and in Northern Ireland. Candidates need to be nominated by past mentees using forms available at [go.nature.com/hmezau](http://go.nature.com/hmezau). Deadline: 4 August. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunq](http://go.nature.com/xhunq)



P. PAPER



## Big Bang blunder bursts the multiverse bubble

*Premature hype over gravitational waves highlights gaping holes in models for the origins and evolution of the Universe, argues Paul Steinhardt.*

When a team of cosmologists announced at a press conference in March that they had detected gravitational waves generated in the first instants after the Big Bang, the origins of the Universe were once again major news. The reported discovery created a worldwide sensation in the scientific community, the media and the public at large (see *Nature* 507, 281–283; 2014).

According to the team at the BICEP2 South Pole telescope, the detection is at the 5–7 sigma level, so there is less than one chance in two million of it being a random occurrence. The results were hailed as proof of the Big Bang inflationary theory and its progeny, the multiverse. Nobel prizes were predicted and scores of theoretical models spawned. The announcement also influenced decisions about academic appointments and the rejections of papers and grants. It even had a role in governmental planning of large-scale projects.

The BICEP2 team identified a twisty (B-mode) pattern in its maps of polarization of the cosmic microwave background, concluding that this was a detection of primordial gravitational waves. Now, serious flaws in the analysis have been revealed that transform the sure detection into no detection. The search for gravitational waves must begin anew. The problem is that other effects, including light scattering from dust and the synchrotron radiation generated by electrons moving around galactic magnetic fields within our own Galaxy, can also produce these twists.

The BICEP2 instrument detects radiation at only one frequency, so cannot distinguish the cosmic contribution from other sources. To do so, the BICEP2 team used measurements of galactic dust collected by the Wilkinson Microwave Anisotropy Probe and Planck satellites, each of which operates over a range of other frequencies. When the BICEP2 team did its analysis, the Planck dust map had not yet been published, so the team extracted data from a preliminary map that had been presented several months earlier. Now a careful reanalysis by scientists at Princeton University and the Institute for Advanced Study, also in Princeton, has concluded that the BICEP2 B-mode pattern could be the result mostly or entirely of foreground effects without any contribution from gravitational waves. Other dust models considered by the BICEP2 team do not change this negative conclusion, the Princeton team showed (R. Flauger, J. C. Hill and D. N. Spergel, preprint at <http://arxiv.org/abs/1405.7351>; 2014).

The sudden reversal should make the scientific community contemplate the implications for the future of cosmology experimentation and theory. The search for gravitational waves is not stymied. At least eight experiments, including BICEP3, the Keck Array and Planck, are already aiming at the same goal.

This time, the teams can be assured that the

world will be paying close attention. This time, acceptance will require measurements over a range of frequencies to discriminate from foreground effects, as well as tests to rule out other sources of confusion. And this time, the announcements should be made after submission to journals and vetting by expert referees. If there must be a press conference, hopefully the scientific community and the media will demand that it is accompanied by a complete set of documents, including details of the systematic analysis and sufficient data to enable objective verification.

The BICEP2 incident has also revealed a truth about inflationary theory. The common view is that it is a highly predictive theory. If that was the case and the detection of gravitational waves was the ‘smoking gun’ proof of inflation, one would think that non-detection means that the

theory fails. Such is the nature of normal science. Yet some proponents of inflation who celebrated the BICEP2 announcement already insist that the theory is equally valid whether or not gravitational waves are detected. How is this possible?

The answer given by proponents is alarming: the inflationary paradigm is so flexible that it is immune to experimental and observational tests. First, inflation is driven by a hypothetical scalar field, the inflaton, which has properties that can be adjusted to produce effectively any outcome. Second, inflation does not end with a universe with uniform properties, but almost inevitably leads to a multiverse with an infinite number of bubbles, in which the cosmic and physical properties vary from bubble to bubble. The part of the multiverse that we observe corresponds to a piece

of just one such bubble. Scanning over all possible bubbles in the multiverse, everything that can physically happen does happen an infinite number of times. No experiment can rule out a theory that allows for all possible outcomes. Hence, the paradigm of inflation is unfalsifiable.

This may seem confusing given the hundreds of theoretical papers on the predictions of this or that inflationary model. What these papers typically fail to acknowledge is that they ignore the multiverse and that, even with this unjustified choice, there exists a spectrum of other models which produce all manner of diverse cosmological outcomes. Taking this into account, it is clear that the inflationary paradigm is fundamentally untestable, and hence scientifically meaningless.

Cosmology is an extraordinary science at an extraordinary time. Advances, including the search for gravitational waves, will continue to be made and it will be exciting to see what is discovered in the coming years. With these future results in hand, the challenge for theorists will be to identify a truly explanatory and predictive scientific paradigm describing the origin, evolution and future of the Universe. ■

**Paul Steinhardt** is professor of physics at Princeton University.  
e-mail: [steinh@princeton.edu](mailto:steinh@princeton.edu)

THE INFLATIONARY  
PARADIGM IS  
FUNDAMENTALLY  
**UNTESTABLE,**  
AND HENCE  
SCIENTIFICALLY  
**MEANINGLESS.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/yhezeh](http://go.nature.com/yhezeh)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## CANCER

### Tumour reined in by its neighbours

Two groups have found that certain non-cancerous cells that surround a tumour in the pancreas keep it in check, even though these cells seem to boost other types of cancer.

Raghu Kalluri at the University of Texas MD Anderson Cancer Center in Houston and his colleagues deleted fibroblast cells, which are found in connective tissue, in a mouse model of pancreatic cancer. These animals had more aggressive tumours and died sooner than those with fibroblasts. These cells seem to help the immune system to fight the tumour. In human pancreatic tissue samples, those with fewer fibroblasts were associated with lower patient survival rates.

Kenneth Olive at Columbia University in New York, Ben Stanger at the University of Pennsylvania in Philadelphia and their colleagues also found that in animals that had fewer connective tissue cells in their pancreatic tumours, the tumours grew faster and had more blood vessels than in control mice.

The results could help to explain why certain drugs that target a tumour's neighbouring cells have failed in pancreatic cancer clinical trials, the authors say.

*Cancer Cell* <http://doi.org/szgj>;  
<http://doi.org/szh> (2014)

## IMMUNOLOGY

### Fetal cells have a good memory

Immune cells that 'remember' past encounters with foreign molecules have been found in human umbilical cord blood, even though the fetal environment



## PALAEONTOLOGY

### Nectar feast in fossil belly

A 47-million-year-old fossil bird with pollen grains in its belly is the first direct evidence of nectar-feeding in birds.

Gerald Mayr and Volker Wilde at the Senckenberg Research Institute in Frankfurt, Germany, analysed the fossil of a small bird, *Pumiliornis tessellatus* (pictured), which is similar to cuckoos or parrots. Its thin, long beak resembles those of other nectar-slurping birds, such as hummingbirds. Electron microscopy revealed pollen grains in the bird's stomach (boxed).

*P. tessellatus* is not closely related to present-day nectar-feeding birds, suggesting that interactions between birds and flowers predate those species, the authors say.

*Biol. Lett.* <http://doi.org/szj> (2014)

is presumed to be sterile.

Human T cells begin to form when the fetus is about ten weeks old, but the cells were thought to remain unexposed

to foreign antigens until after

birth. However, Richard Lo-Man of the Pasteur Institute in Paris and his team found a subset of fetal T cells known as

memory cells, which remember foreign molecules and respond more rapidly to them during a second exposure. These memory cells make up 1–6% of CD4, or helper, T cells in healthy neonatal cord blood. The identities of the antigens they react with are unknown.

*Sci. Transl. Med.* 6, 238ra72 (2014)

## PARTICLE PHYSICS

### Antimatter passes charge test

Physicists have tightened the limits on the possible charge of an atom of antihydrogen.

The neutral charge on many atoms and molecules has been measured with extremely high precision. The standard model of physics says that hydrogen's antimatter counterpart should have an opposite charge and so be neutral to a similar level. Any differences between the two could help to explain why the Universe contains more matter than antimatter.

Joel Fajans at the University of California, Berkeley, and his colleagues used data from previous experiments to analyse the influence of electric fields on antihydrogen atoms released from a magnetic trap. They found that the atom was charge-neutral, with a limit 1 million times lower than the best previous figure.

*Nature Commun.* <http://dx.doi.org/10.1038/ncomms4955> (2014)

## BIOGEOGRAPHY

### Ancient coral reefs protected fish

As the climate fluctuated over the past 3 million years, stable coral reefs provided a safe haven for tropical fish species, leading to the vast diversity of reef fishes seen today.

GERALD MAYR/SENCKENBERG RES. INST.



David Mouillot at Montpellier University in France and his colleagues used sediment cores to map coral reef habitats over the past 3 million years and compared this analysis with the current distribution of more than 6,000 reef fish species. The team found that the historical proximity of the fish habitats to stable coral reefs during cold periods had a greater influence on current patterns of reef fish diversity than present-day environmental factors such as sea-surface temperature. The diversity of damselfishes (*Pomacentridae*, pictured) in particular was strongly linked to their distance from coral reef refuges.

As the climate warms, conservation biologists should focus on protecting coral reefs that connect historical refuges.

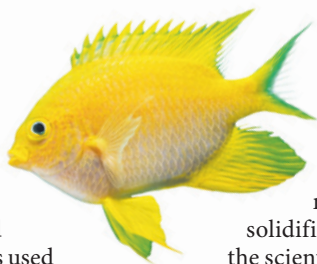
*Science* 344, 1016–1019 (2014)

## GEOLOGY

## First crust looked like Iceland

Ancient rocks from Canada suggest that Earth's first land was made billions of years ago in the same way as Iceland.

Geologists think that continental crust first formed as magma broke through the dense ocean crust and crystallized on top. Iceland is considered to be the best modern analogue of early continental crust formation, but until now no ancient rocks have been found with the same chemistry as Icelandic ones. Jesse Reimink and his colleagues at the University of Alberta in Edmonton, Canada, discovered that 4.02-billion-year-old rocks in northwestern Canada share the same geochemical signatures as rocks from Iceland, but have signatures that are distinct from 3.9-billion- to 2.5-billion-year-old rocks.



The Canadian rocks were probably formed as magma cooled and solidified in shallow waters, the scientists say.

*Nature Geosci.* <http://doi.org/szb> (2014)

## NEUROPROSTHETICS

## Brain signals move paralysed hand

Paralysed monkeys were able to move their hands after researchers rerouted brain signals into the animals' spinal cord.

Jonas Zimmermann and Andrew Jackson at Newcastle University, UK, implanted recording electrodes in the premotor cortex area of the brains of two monkeys and stimulating electrodes in the monkeys' spinal cords. The authors trained the animals to pull a lever for a food reward and then temporarily paralysed them by injecting a chemical into the motor cortex.

When the paralysed monkeys tried to pull the lever, a computer program translated the electrical impulses from the brain into stimulation signals that were transmitted to the spinal cord. The animals could move the lever a greater distance and more frequently when the signals were transmitted compared with when they were not.

The approach could one day be used to make paralysed human limbs perform a variety of tasks, the authors say.

*Front. Neurosci.* 8, 87 (2014)

## ATMOSPHERIC SCIENCE

## Climate concerns from refrigerants

Chemicals that are slowly escaping from refrigerators and air conditioners are a greater threat to the climate than previously thought.

Hydrofluorocarbons (HFCs) are refrigerants that are replacing banned ozone-depleting chemicals. But HFCs are also greenhouse gases, and the amount of

## SOCIAL SELECTION

Popular articles on social media

## Winner takes all in science

Many researchers on Twitter have been sharing an article that supports their hard-won observations: science funding doesn't always go to the most deserving. The article, by sociologist Yu Xie at the University of Michigan in Ann Arbor, notes that science now resembles a "winner-takes-all" system that awards a disproportionate share of resources to a minority of researchers and institutions. Timothy O'Leary, a neuroscientist at Brandeis University in Waltham, Massachusetts, tweeted that "inequality is OK provided the 90% can make a living". But even that is proving difficult. "Postdocs are having permanent conversations about making the transition out of science," he says. "They have to move up or move out."

Xie, Y. *Science* 344, 809–810 (2014)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**  
For more on popular papers:  
[go.nature.com/f79a6z](http://go.nature.com/f79a6z)

these chemicals stored in refrigeration equipment is increasing. Guus Velders at the National Institute for Public Health and the Environment in Bilthoven, the Netherlands, and his colleagues analysed different scenarios for reducing HFCs, either by phasing out production or by destroying HFC-containing equipment.

They found that phasing out production of HFCs early, such as by 2020, would provide the greatest environmental benefit. If the phase-out were to occur later, millions of refrigeration and air-conditioning units would need to be destroyed to achieve the same benefit.

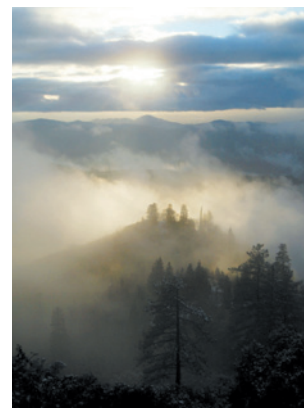
*Atmos. Chem. Phys.* 14, 4563–4572 (2014)

## CLIMATE CHANGE

## Fading winter fog threatens crops

Climate change has reduced the amount of winter fog in California (pictured), possibly damaging the region's fruit and nut crops.

Much of the fruit and nut production in the United States comes from the Central Valley of California. Dennis



Baldocchi and Eric Waller at the University of California, Berkeley, analysed 33 years of satellite data and found that the number of winter fog events in this region decreased by 46% on average over 32 winters.

Winters with less fog will yield fewer hours of winter chill, which fruit and nut trees need in order to rest and prepare for the next growing season.

*Geophys. Res. Lett.* <http://doi.org/szk> (2014)

➔ **NATURE.COM**  
For the latest research published by Nature visit:  
[www.nature.com/latestresearch](http://www.nature.com/latestresearch)



# SEVEN DAYS

The news in brief

## RESEARCH

### Misconduct update

Haruko Obokata, a stem-cell researcher at the RIKEN Center for Developmental Biology in Kobe, Japan, has agreed to retract one of her two controversial research papers, according to media reports in Japan on 28 May. In the papers published in *Nature* in January, Obokata claimed that she had created a new type of stem cell using simple exposure to acid or physical stress. But the work was found to have included manipulated and duplicated images, and in April a RIKEN investigation found Obokata guilty of misconduct. See [go.nature.com/wsfox5](http://go.nature.com/wsfox5) for more.

### Weight issues

Obesity is on the rise throughout the world, researchers announced on 28 May. Data from more than 19,000 individuals in 188 countries show that the prevalence of overweight and obese people rose by 27.5% for adults and by 47.1% for children between 1980 and 2013. Almost two-thirds of obese people live in developing countries. The study found no countries in which obesity rates had fallen significantly (M. Ng *et al.* *Lancet* <http://doi.org/szv>; 2014).

## POLICY

### IQ on trial

Some defendants with intelligence-quotient (IQ) scores of more than 70 may now escape execution on the basis of intellectual disability, after a US Supreme Court ruling on 27 May. The ruling came in favour of Freddie Lee Hall, a convicted murderer in Florida who has spent 35 years on death row. Federal law prohibits the execution of

intellectually disabled people, defined under Florida law as an IQ of less than 70. In his latest test, Hall had an IQ of 71. The Supreme Court affirmed the position of psychiatrists that IQ tests have a ten-point margin of error. The case now returns to Florida's court for a reassessment. See [go.nature.com/xiiitc](http://go.nature.com/xiiitc) for more.

### EU clinical trials

New rules governing clinical trials across Europe became law on 27 May when they were published in the *Official Journal of the European Union*. The rules were overhauled to simplify how trials are approved, including

harmonizing application procedures across all 28 member states, and include an obligation to disclose research results. They will come into force in mid-2016.

### Patent clarity

The US Supreme Court moved to rein in ambiguous patents in a ruling on 2 June. The decision relates to a case concerning a patent on heart-rate monitors used during exercise. A lower court had ruled that the patent was valid despite being somewhat ambiguous. The Supreme Court said that the lower court should have higher standards for patent clarity. The move

as the Bruce Mansfield Plant in Pennsylvania, pictured) are the single largest source of carbon pollution in the country. Regulations exist to limit emissions of pollutants such as arsenic and mercury, but there are no national controls on carbon-emission levels. The EPA expects to finalize the plan by June 2015, after a 120-day period for public comment. See [go.nature.com/gjyruv](http://go.nature.com/gjyruv) for more.



ROBERT NICKELBERG/GETTY IMAGES

## Cleaner power in prospect

could have implications for 'patent trolls', who take advantage of ambiguous patents to sue companies for infringement. See page 7 for more.

### Rights to therapy

Patients do not have an automatic right to try a compassionate therapy when there is no scientific evidence that it works, according to a landmark ruling of the European Court of Human Rights in Strasbourg, France. The 28 May ruling referred to the case of Nivio Durisotto, who wanted his daughter, who has a degenerative brain disease, to be treated with

a controversial stem-cell therapy offered by the Stamina Foundation in Brescia, Italy. Stamina's president Davide Vannoni — who is facing charges of fraud — had encouraged patients to appeal for the right to treatment on compassionate grounds. See [go.nature.com/xqdpfb](http://go.nature.com/xqdpfb) for more.

## GM compromise

The European Union (EU) reached a compromise on the cultivation of genetically modified (GM) crops, in a closed-door meeting on 28 May. EU member states are currently supposed to allow cultivation of any GM crop that the European Food Safety Authority deems safe for health and the environment. But some countries have refused, creating a constitutional stand-off. Now, EU member states will be allowed to opt out of growing GM crops on non-scientific grounds, such as issues relating to town and country planning. European environment ministers will vote on the proposal on 12 June; it also requires parliamentary approval.

## AWARDS

### Kavli prizes

Nine scientists shared the three 2014 biennial Kavli prizes (worth US\$1 million



each), announced on 29 May. Alan Guth, Andrei Linde and Alexei Starobinsky won in astrophysics, for their work developing the theory of cosmic inflation. The nanoscience prize went to Thomas Ebbsen, Stefan Hell and John Pendry, for studies defeating the 'resolution limit' of optical imaging and microscopy — showing that light can interact with nanostructures smaller than the light's wavelength. In neuroscience, Brenda Milner (**pictured**), John O'Keefe and Marcus Raichle won for pinpointing brain regions involved in memory and the key functions of specialized nerve cells. See [go.nature.com/ifevbf](http://go.nature.com/ifevbf) for more.

### Shaw prizewinners

The Shaw Prize Foundation in Hong Kong announced the winners of its annual prizes on 27 May. Worth US\$1 million, the prizes

are awarded in the areas of astronomy; mathematical sciences; and life science and medicine. The winners include Daniel Eisenstein of Harvard University in Cambridge, Massachusetts, for his work on waves originating in the early Universe called baryonic acoustic oscillations, and Kazutoshi Mori of Kyoto University in Japan for his part in discovering a key mechanism involved in degenerative diseases. See [go.nature.com/vrpwla](http://go.nature.com/vrpwla) for more.

## FUNDING

### Stem-cell petition

The European Commission has rejected a petition for legislation to ban the funding of research that uses human embryonic stem cells. The petition, signed by 1.7 million European citizens, came from a pro-life organization based in Brussels called One of Us. On 28 May, the commission said that stem-cell research held great promise and it was therefore in the public interest to support it. It also agreed that human embryonic stem cells are still sometimes required in such research. See [go.nature.com/aa6tuv](http://go.nature.com/aa6tuv) for more.

### Brain grants

The first round of awards from a large US multi-agency funding effort to study the

## COMING UP

### 8–13 JUNE

Geoscientists will meet at Goldschmidt2014 in Sacramento, California. Up for discussion are the potential habitability of Mars, and the environmental legacy of California's gold rush. See [go.nature.com/o8iznc](http://go.nature.com/o8iznc)

brain have gone to work on neuropsychological disorders. On 27 May, the Defense Advanced Research Projects Agency in Arlington, Virginia, awarded the University of California, San Francisco, US\$26 million and Massachusetts General Hospital in Boston \$30 million over five years. The projects are funded under the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) initiative. They will focus on creating devices that stimulate areas deep in the brain and record brain activity for research on soldiers and veterans with conditions such as post-traumatic stress disorder.

## BUSINESS

### Sequencing boost

Swiss drug company Roche, based in Basel, moved to boost its genetic-sequencing business by acquiring Genia Technologies of Mountain View, California. Roche announced on 2 June that it will pay up to US\$350 million for Genia, which is developing a semiconductor-based DNA-sequencing technology that uses nanopores to read single genetic molecules. Roche wants to use sequencing in its diagnostics business. In 2012, it unsuccessfully tried to buy Illumina of San Diego, California, which owns most of the global sequencing market.

► **NATURE.COM**

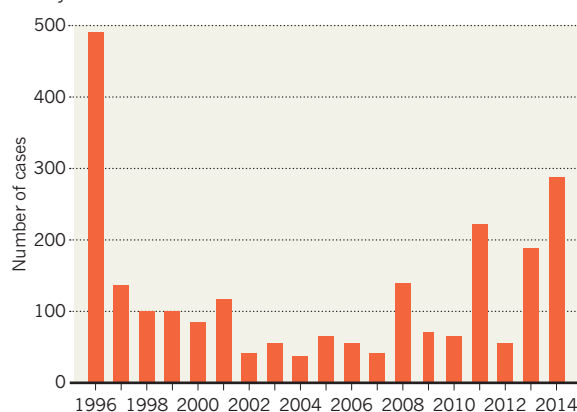
For daily news updates see: [www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

This year, the United States has already experienced the greatest number of measles cases for 20 years, researchers from the Centers for Disease Control and Prevention in Atlanta, Georgia, announced last week. The country declared the disease eliminated in 2000, and more than 90% of its young population are vaccinated. The most likely reason for the rise, the researchers say, is that the virus is being imported by unvaccinated US travellers who then mingle with other unvaccinated people.

### MEASLES SPIKE

US public-health experts urge vaccination after largest outbreak in 20 years.





# NEWS IN FOCUS

**CLIMATE CHANGE** Grasslands experiments aim to fix holes in vegetation knowledge **p.16**

**REGENERATIVE MEDICINE** Salvaged stem-cell trial offers fresh hope for spinal injuries **p.18**

**POLICY** Constitutional changes free up German science funding **p.19**



**ENERGY** Race to build an artificial leaf in the lab heats up **p.22**

ELIAVA INST.



Bacteriophages could be a resource for fighting drug-resistant bacterial infections.

## MICROBIOLOGY

# Phage therapy gets revitalized

*The rise of antibiotic resistance rekindles interest in a century-old virus treatment.*

BY SARA REARDON

For decades, patients behind the Iron Curtain were denied access to some of the best antibiotics developed in the West. To make do, the Soviet Union invested heavily in the use of bacteriophages — viruses that kill bacteria — to treat infections. Phage therapy is still widely used in Russia, Georgia and Poland, but never took off elsewhere. “This is a virus, and people are afraid of viruses,” says Mzia Kutateladze, who is the head of the scientific council at the Eliava Institute in Tbilisi, which has been studying phages and using

them to treat patients for nearly a century.

Now, faced with the looming spectre of antibiotic resistance, Western researchers and governments are giving phages a serious look. In March, the US National Institute of Allergy and Infectious Diseases listed phage therapy as one of seven prongs in its plan to combat antibiotic resistance. And at the American Society for Microbiology (ASM) meeting in Boston last month, Grégory Resch of the University of Lausanne in Switzerland presented plans for

Phagoburn: the first large, multi-centre clinical trial of phage therapy for human infections, funded by the European Commission.

Ryland Young, a virologist at Texas A&M University in College Station, attributes the previous lack of Western interest to clinicians’ preference for treating unknown infections with broad-spectrum antibiotics that kill many types of bacterium. Phages, by contrast, kill just one species or strain. But researchers now realize that they need more precise ways to target pathogenic bacteria, says microbiologist Michael Schmidt of the Medical University of South Carolina in Charleston. Along with the rising tide of strains resistant to last-resort antibiotics, there is growing appreciation that wiping out the human body’s beneficial microbes along with disease-causing ones can create a niche in which antibiotic-resistant bacteria can thrive. “Antibiotics are a big hammer,” Schmidt says. “You want a guided missile.”

Finding a phage for a bacterial target is relatively easy, Young says. Nature provides an almost inexhaustible supply: no two identical phages have ever been found. As a bacterium becomes resistant to one phage — by shedding the receptor on the cell surface that the virus uses to enter — the Eliava Institute researchers simply add more phages to the viral cocktails that patients receive. Kutateladze says that they update their products every eight months or so, and do not always know the exact combination of phages that make up the cocktail.

Resch, who is one of the leaders of the Phagoburn study, says that regulatory agencies would need to figure out how to oversee such a rapidly evolving product before the therapy could progress beyond clinical trials. He hopes that phage therapy will be treated in a similar way to the seasonal influenza vaccine, for instance, which is updated every year as new flu strains emerge.

The fact that the European Union (EU) is contributing €3.8 million (US\$5.2 million) to the Phagoburn study shows that it is willing to consider the approach, Resch says. Beginning in September, researchers in France, Belgium and the Netherlands plan to recruit 220 burn victims whose wounds have become infected with the common bacteria *Escherichia coli* or *Pseudomonas aeruginosa*. The patients will be given phage preparations from a company in Romainville, France, called Pherecydes Pharma, which has isolated more than 1,000 viruses from sources such as sewage or river water and screened them for the ability to kill ►

**NATURE.COM**  
For more on  
antibiotics and  
phage therapy, see:  
[go.nature.com/pykv3d](http://go.nature.com/pykv3d)



► pathogenic bacteria. To lower the chance that resistance will develop, the patients will receive a cocktail of more than a dozen phages that enter bacterial cells in different ways. If the phage treatment fails, patients will then receive standard antibiotics.

Although governments are starting to pay attention to phage therapy, pharmaceutical companies remain reluctant to get on board, Young says. Because phage therapy is nearly a century old, it would be difficult for a company to claim a treatment as intellectual property, and therefore recoup its costs. Young says it is likely that a 2013 ruling by the US Supreme Court against the patenting of natural genes would also apply to phages isolated from nature. Jérôme Gabard, chief executive of Pherecydes, says that the company is banking on hopes that developing and characterizing precise combinations of natural phages to target particular bacteria will be patentable.

An engineered phage could, in theory, be patented. At the ASM meeting last month, researchers led by synthetic biologist Timothy Lu at the Massachusetts Institute of Technology in Cambridge presented work on a phage engineered to use a DNA-editing system called CRISPR to kill only antibiotic-resistant bacteria. The phage injects the bacterium with DNA, which the microbe transcribes into RNA. If part of the bacterium's antibiotic-resistance gene matches that RNA sequence, an enzyme called Cas9 cuts up the cell's DNA, killing it.

In initial trials, the researchers found that their phage could kill more than 99% of the *E. coli* cells that contained specific antibiotic-resistance gene sequences, whereas it left susceptible cells alone. Giving the phage to waxworm larvae infected with resistant *E. coli* increased the worms' chance of survival. The researchers are now starting to test the system in mice (human trials are a long way off).

Gabard does not expect that phage therapy will ever replace antibiotics. But he says that he can envisage regulatory agencies approving it for patients in whom drug treatments have failed. And some people with antibiotic-resistant infections are taking matters into their own hands. Kutateladze says that an increasing number of EU patients are travelling to Georgia for phage treatment. She adds that doctors in some EU countries send patients' samples to the Eliava Institute, which then sends back a phage cocktail specific to the bacterium causing the infection. "When there's no hope, you'll do anything," Schmidt says.

Meanwhile, researchers are watching the Phagoburn study with interest, hoping that it will lay the groundwork for moving the technology into the clinic. "We just need one really great success for the field to really open up," says Lu. ■



Sections of Mongolian grassland will undergo climate-manipulation analysis at the Duolun research centre.

#### GLOBAL WARMING

# Land models put to climate test

*Study under way on Mongolian steppes aims to improve knowledge of warming effects on vegetation.*

BY JANE QIU

Shiqiang Wan remembers his first experience of a major sandstorm more than a decade ago in Inner Mongolia. "It was like sands being dumped on me by a gigantic dustbin," he says. "I couldn't see anything just a few metres away." Decades of overgrazing were turning the region into desert.

Grassland is now developing there again after strict grazing limits were imposed on the autonomous region by China in 2000. But Wan, an ecologist at Henan University in Kaifeng, worries that another, more challenging, menace will eat away at vegetation

there and elsewhere: climate change. And if grasslands wither on a global scale, "it would not only cause widespread desertification, but accelerate climate warming by increasing carbon dioxide levels in the air", he says.

Yet scientists know little about the effects of climate change on land ecosystems, or how they will affect atmospheric CO<sub>2</sub>.

Wan and his colleagues have an ambitious plan for filling that information vacuum. In a US\$260,000 operation, they have cut 54 6-tonne chunks of soil — 2.2 metres by 1.5 metres, by 1.2 metres deep — from three types of grassland in the Mongolian Plateau, and are now growing them at the

SHIQIANG WAN

SOURCE: SHIQIANG WAN &amp; SHILONG PIAO

Duolun Ecological Research Station in the region. Later this month, the team will start a decade-long experiment to test how artificially induced warming and varying amounts of rainfall affect vegetation growth, water, nutrients and plant composition. More crucially, it will investigate the underlying mechanisms.

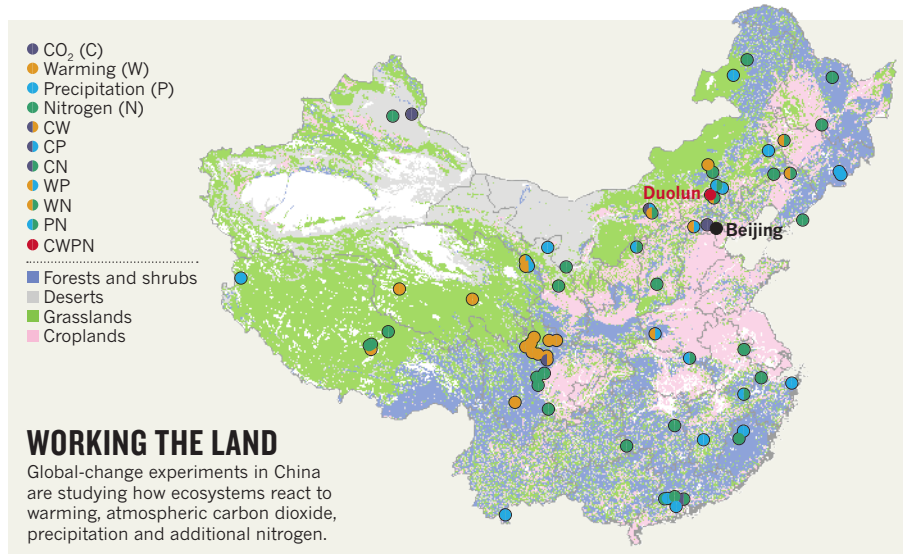
The importance of the experiment “goes beyond the Mongolian steppes”, says Alan Knapp, an ecologist at Colorado State University in Fort Collins. “Grasslands across broad temperature and precipitation gradients can respond quite differently to climate change — in some cases, as differently as between a grassland and a forest.”

Models that simulate how terrestrial ecosystems respond to climate change and affect atmospheric CO<sub>2</sub> levels fail to capture these nuances, Knapp told a meeting last month in Beijing. The meeting was funded by the US National Science Foundation with the aim of fostering closer ties between experimentalists and modellers. The models used assume that a given type of vegetation will behave in the same way, regardless of the local climate. “Historically, ecologists tend to work in their backyard and assume everywhere else is the same,” Knapp says.

Not only that, but studies conducted at different locations are often not comparable because different teams address different questions, and their experimental protocols also vary. For this and other reasons, few data from the hundreds of manipulation experiments conducted worldwide — at a cost of billions of dollars — have been incorporated into the models, says Yiqi Luo, a climate modeller at the University of Oklahoma in Norman.

Until recently, “there has been limited cross-over between experimentalists and modellers”, says Jeff Dukes, an ecologist at Purdue University in West Lafayette, Indiana. Typically, “we would turn to modellers years after an experiment has begun and then realize things that we should have measured are not measured”, adds Richard Norby, an ecologist at Oak Ridge National Laboratory in Tennessee. “We would have made a lot faster progress if the two communities had worked closely together from the very beginning.”

The key to success lies partly in collecting the right data for the model from the start. Wan’s study will be among the first to do this, and is one of a growing number of



such ‘global-change’ experiments for land ecosystems (see ‘Working the land’). It will measure a range of factors that are needed to drive models, such as soil carbon content and nutrient level, plant biomass and productivity, and meteorological parameters. Results from the study could, for example, be the first step to making comparisons across a rainfall gradient, and so improve land models, Luo says.

Wan’s initiative will be welcome. The performance of land models “is rather poor at the moment”, says Philippe Ciais, a climate modeller at the National Centre for Scientific Research in Paris. For instance, their predictions of CO<sub>2</sub>

**“Grasslands across broad temperature and precipitation gradients can respond quite differently to climate change.”**

release from land ecosystems in response to changes in atmospheric CO<sub>2</sub>, temperature and other climate variables diverge wildly (V. K. Arora *et al. J. Clim.* **26**, 5289–5314; 2013).

Land models “are the weakest link in our projection of future climate change”, Ciais adds.

To understand why the models differ so much, the US Department of Energy is funding a project, led by Norby, to compare experimental data and land models. “Initially, we wanted to see which models fit the data best,” says Norby. “But we soon realized that this isn’t the right approach because models can give the

right answer for the wrong reason.”

The team found, for instance, that most models are flawed in their basic assumptions about how plants use nitrogen, but can still reproduce some of the effects seen in global-change experiments because errors in opposite directions cancel each other out (S. Zaehle *et al. New Phytol.* <http://dx.doi.org/10.1111/nph.12697>; 2014). “This means that they are not going to be very powerful for predicting the future,” says Norby.

At the Beijing meeting, researchers identified other weaknesses. The models are poor at capturing how vegetation composition responds to climate change — a crucial factor for predicting changes in carbon storage in the long term. This is partly because the models tend to be based on the physiology of single plants and partly because they take scant account of plant diversity.

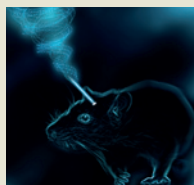
To improve land models, researchers at the meeting plan to launch long-term initiatives to probe how ecosystems worldwide respond to climate change and to standardize experimental designs and methods so that the results can be compared.

The ecologists are also calling for a common repository for data from global-change experiments that can be accessed by modellers — similar to existing databases for atmospheric and oceanic observations.

“There is a long way to go,” says Luo. “To connect experimentalists and modellers is the key. The bridge has just started being built.” ■

**MORE ONLINE**

#### TOP NEWS



Optogenetics provides direct evidence for memory model [go.nature.com/zsprrs](http://go.nature.com/zsprrs)

#### MORE NEWS

- Publication record predicts who will become a principal investigator [go.nature.com/ledkav](http://go.nature.com/ledkav)
- Lianas protect trees from lightning [go.nature.com/rchaaz](http://go.nature.com/rchaaz)
- Stellar winds make planets uninhabitable [go.nature.com/i7qtkp](http://go.nature.com/i7qtkp)

#### NATURE PODCAST



Malnutrition and microbiota; the basis of memory; and China moves mountains to build cities [nature.com/nature/podcast](http://nature.com/nature/podcast)

SADEGH NABAVI &amp; SINA ALIZADEH



# Funding windfall rescues abandoned stem-cell trial

*But more players have joined the quest to treat spinal-cord injury with embryonic cells.*

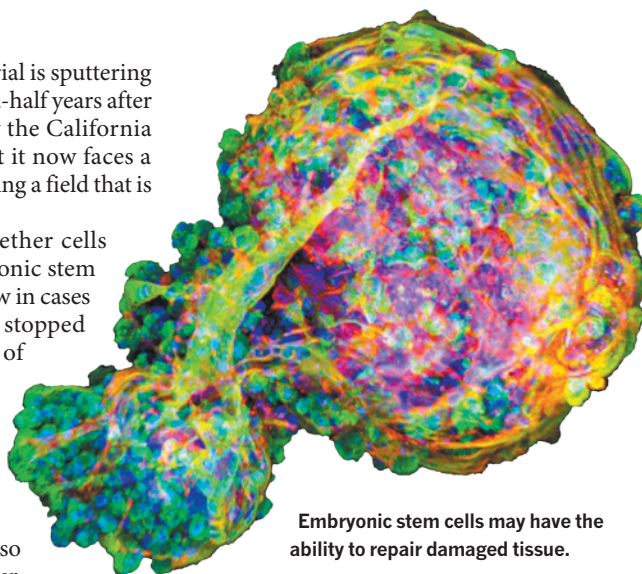
BY ERIKA CHECK HAYDEN

A landmark stem-cell trial is sputtering back to life two-and-a-half years after it was abandoned by the California company that started it. But it now faces a fresh set of challenges, including a field that is packed with competitors.

The trial aims to test whether cells derived from human embryonic stem cells can help nerves to regrow in cases of spinal-cord injury. It was stopped abruptly in 2011 by Geron of Menlo Park, California (see *Nature* 479, 459; 2011); the firm said at the time that it wanted to focus on several promising cancer treatments instead. Now, a new company — Asterias Biotherapeutics, also of Menlo Park — plans to resurrect the trial with a US\$14.3-million grant that it received on 29 May from the California Institute for Regenerative Medicine (CIRM), the state's stem-cell-funding agency.

But the field has moved on since Geron treated its first patient in 2010, and the therapy that Asterias inherited is no longer the only possibility for spinal-cord injury. StemCells, a biotechnology company in Newark, California, has treated 12 patients in a safety study of a different type of stem cell, and it plans to start a more advanced trial this year to test effectiveness. And another entrant to the field, Neuralstem of Germantown, Maryland, received regulatory approval in January 2013 to begin human tests of its stem-cell product.

Geron's human trial was the first approved to use cells derived from human embryonic stem cells. But regulators halted it twice, once citing concerns about the purity and predictability of the cells being implanted, and again after the company reported seeing microscopic cysts in the spinal cords of rats that had been treated in preclinical studies. The worry was that the cysts could be teratomas — uncontrolled growths that can form from embryonic stem cells, a feared side effect of treatment. Geron later said that the growths were not teratomas, and the US Food and Drug Administration allowed the trial to proceed. But after injecting the cells into five of the ten intended patients, the company said that it had run out of money for the trial.



Embryonic stem cells may have the ability to repair damaged tissue.

Geron founder Michael West and former chief executive Thomas Okarma then formed Asterias, which bought Geron's stem-cell therapy last year. The company plans first to treat three patients with spinal-cord damage in the neck, using a low dose of the stem cells; it will then treat different people with higher doses to see if the therapy can restore any sensation or function in the trunk or limbs.

The five patients previously treated by Geron, whom Asterias continues to track, had cord damage at chest level. On 22 May, Asterias reported that none of those five had experienced serious side effects from the treatment or developed immune responses to it.

Researchers say that the continuation of the former Geron trial is important because it uses a type of cell different from the fetus-derived ones used by StemCells and Neuralstem. Geron surgically implanted embryonic stem cells that had been coaxed *in vitro* to grow into immature myelinated glial cells, which insulate nerve fibres when mature. The other companies are using partially differentiated cells derived from fetal brain tissue, which might produce substances that protect surviving tissue and make new connections in the neural circuitry.

"It's very good for the field, because we now have multiple cell lines being tested in very similar populations of patients, and this will help us define what is needed to make this approach work," says Martin Marsala, a neuroscientist

at the University of California, San Diego, whose work has shown that Neuralstem's cells can develop into working neurons and restore movement to rats with cord injuries in the neck.

Still, some in the field question whether stem cells are a good candidate for treating paralysis. Spinal trauma can destroy a range of cells; it can be difficult to access the injury site; and damage can be very severe. Treating paralysis "may be asking a lot more of the cells than they are capable of biologically", says Robert Lanza, chief scientific officer at Advanced Cell Technology in Marlborough, Massachusetts, which is now testing embryo-derived stem cells in two eye diseases.

Lanza says that his company opted not to buy Geron's stem-cell therapy in part because the treatment did not seem effective enough in animal models. Some of the scientists who reviewed Asterias's grant proposal to the CIRM also expressed concern about the treatment's effectiveness in animals. "It would be fantastic if [the cells] work, but I think it would also be wise not to set expectations too high," says Lanza.

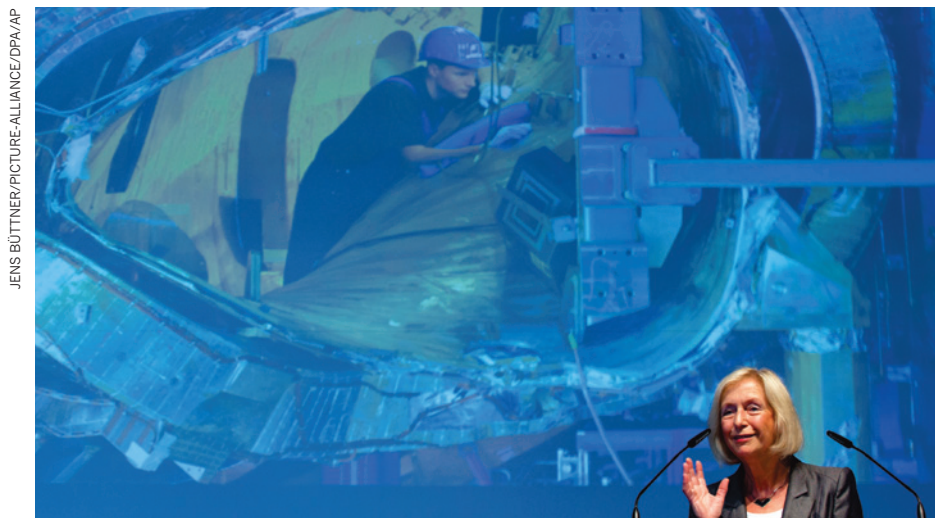
Marsala counters that the cells do not have to heal the spinal cord completely to improve people's lives. He notes that simply replacing injured or dead cells with healthy ones might prevent the damage from spreading. StemCells has reported that four people in its studies have recovered some sensation; two previously had no feeling in their lower limbs and another two had some but were able to feel more after the cell transplant. Marsala sees this as a sign that the therapy is working. "Nobody is claiming that you are going to see complete recovery in these patients," he says.

Christopher Scott, an ethicist at Stanford University in California, says that before scientists can judge the effectiveness of the therapies, the companies must complete their trials. But he wonders whether Asterias has sufficient resources to make good on its promise to carry Geron's therapy through to completion.

Scott notes that Geron spent hundreds of millions of dollars developing spinal-cord and other embryonic-stem-cell therapies, but managed to treat only five people. The restarted trial will face even greater scrutiny, and he worries that if it were halted, it would be a huge blow to patients. "Bioethicists and policy analysts will be watching carefully," he says. ■

NIBSC/SPL





Research minister Johanna Wanka says planned changes will improve German universities.

## POLICY

# Federal boost for German science

*Agreement on changes to constitution pave way for university funding shake-up.*

BY QUIRIN SCHIERMEIER

Germany's highly federalized political system has been a pillar of democracy for more than half a century. But even as the country's constitution has underpinned its stability, it has been a thorn in the side of the scientific community. Universities are largely funded by local governments, and there are tight restrictions on the uses of federal funding — for example, it cannot be used to pay permanent staff or infrastructure costs. This makes it difficult for institutions to compete with the world's leading research powerhouses.

But now, after years of discussions, German legislators have finally agreed to revise a paragraph in the constitution that prevents the federal government from providing permanent direct funding to universities. The move has pleased scientists and promises to pave the way for improved job security, lucrative research collaborations and possibly new centres.

Education has traditionally been the responsibility of Germany's 16 states, and they have been keen to maintain control. In 2006, for example, they signed up to a reform that abolished federal interference in higher education and increased their independence. This made it even more difficult for the federal government to contribute to the maintenance of the more than 100 research universities run by the states.

There are exceptions: the federal government provides more than half of the €2.5-billion (US\$3.4-billion) annual budget of the DFG, Germany's main grant-giving agency for university research (the rest comes mainly from the states). The federal science ministry also contributes 75% of a €4.6-billion excellence initiative, launched in 2006, that allows universities to compete for top-up funds to bolster neglected disciplines. Since its inception, almost 100 graduate schools and large

research collaborations have been set up at 39 universities with this money. In total, the federal government pumps about €17 billion into the annual research budget, a figure dwarfed by the states' €75 billion. Germany's total research and development budget is roughly 2.9% of gross domestic product (see 'Big spenders').

But a lack of permanent federal support creates problems. For example, the excellence initiative expires in 2017, prompting fears that projects will be discontinued and researchers laid off. Scientists therefore applauded the planned constitutional change, which will allow the federal government to step in with permanent financial support for projects at risk.

"We now have much more room to manoeuvre," says Johanna Wanka, federal minister for education and research. "This is a great political success and it will lastingly improve conditions at German universities."

The changes, agreed last week and expected to come into effect before 2015, will end a six-month impasse over a €6-billion fund that the government last year promised to inject into science and education between 2014 and 2017. They will also enable a fundamental reform of German science, the heads of Germany's large research organizations said in a joint statement.

"The promised constitutional change is a true sea change that will finally enable the government to directly and permanently co-finance universities," says Wolfgang Marquardt, head of the German Science Council, which advises the government on science policy.

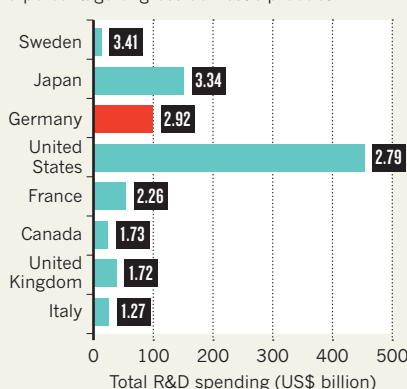
Universities in Germany's poorer states, such as the University of Bremen, a past winner of an excellence initiative award, are particularly in need of a steady flow of federal money. Gerold Wefer, a marine scientist at MARUM, the university's centre for marine environmental sciences, which was set up with award funds, is now optimistic that the centre can survive beyond 2017. "Otherwise, dozens of scientists on fixed-term contracts might have to fear for their jobs," he says.

How any future cash from the federal government will be used is yet to be worked out. The government has previously ruled out continuing the excellence initiative in its current form. Instead, says Marquardt, the government could channel extra cash to the DFG, enabling the agency to fund existing or new collaborations on a competitive basis. Alternatively, he says, the new money could be used to create federally funded university institutes or regional science centres to continue and expand successful research collaborations. For example, Wolfgang Herrmann, president of the Technical University of Munich and a long-time champion of a German national institute of technology, has suggested that his university could become a federal technology hub.

But Marquardt cautions that it is early days. "The changes agreed upon open up many pleasant possibilities," he says. "We shall now start thinking about what might work best." ■

## BIG SPENDERS

Total research and development (R&D) spend in 2012 by selected countries, and the amounts as a percentage of gross domestic product.



SOURCE: OECD

# Big Bang finding challenged

*Signal of gravitational waves was too weak to be significant, studies suggest.*

BY RON COWEN

The astronomers who earlier this year announced that they had evidence of primordial gravitational waves jumped the gun, two independent analyses suggest.

The papers, published on the arXiv preprint repository, propose that the original analysis did not properly account for the confounding effects of galactic dust. Although further observations may yet confirm the findings, independent researchers now say they no longer think that the original data constituted significant evidence.

"Based on what we know right now, we have no evidence for or against gravitational waves," says Uroš Seljak, a cosmologist at the University of California, Berkeley, and co-author of one of the latest studies<sup>1</sup>.

Astronomers using the BICEP2 radio telescope at the South Pole announced in March that they had found a faint twisting pattern in the polarization of the cosmic microwave background (CMB), the radiation left over from the Big Bang. This pattern, they said, was evidence for primordial gravitational waves — ripples in the fabric of space-time generated in the Universe's first moments. The findings were widely hailed as confirmation of the theory of cosmic inflation, which holds that the cosmos ballooned in size during the first fraction of a second after the Big Bang.

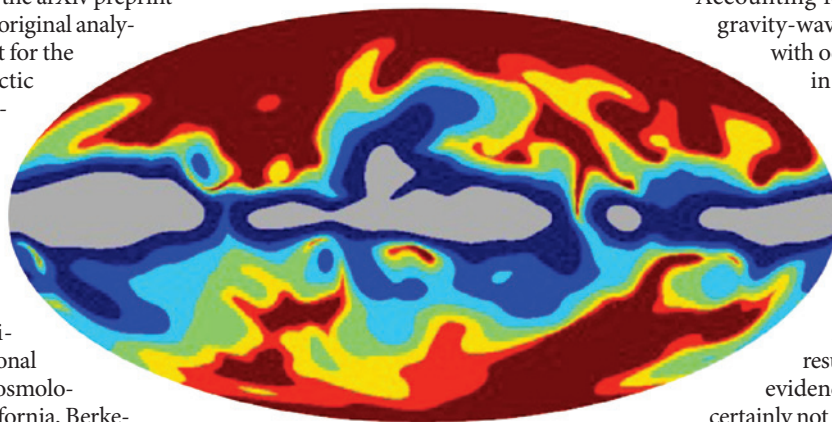
But the new analyses suggest that the twisting patterns in the CMB polarization could just as easily be accounted for by dust in the Milky Way<sup>1,2</sup>.

The papers follow a presentation three weeks ago by Raphael Flauger, a theoretical physicist at New York University and the Institute for Advanced Study in Princeton, who re-examined a map of galactic dust used by BICEP2. Flauger concluded that the BICEP2 researchers had probably underestimated the fraction of polarization caused by dust in the map, which was compiled from data from the European Space Agency's Planck spacecraft. Flauger says that when the dust is fully accounted for, the signal that can be attributed to gravitational waves either vanishes or is greatly diminished.

"I had thought that the [BICEP2] result was very secure," said Alan Guth, the cosmologist who first proposed the inflation concept in 1980, after learning about Flauger's talk.

"Now the situation has changed," added Guth, who works at the Massachusetts Institute of Technology in Cambridge.

The BICEP2 researchers have argued that the Planck map figured in only one of the six models that they used to examine the role of



Cross-checks of gravitational-wave results used this galactic-dust map from the Planck satellite.

dust. But in a paper<sup>2</sup> posted to the arXiv server on 28 May, Flauger and his co-authors David Spergel and Colin Hill, both of Princeton University in New Jersey, say that the five other models are based on a low estimate — between 3.5% and 5% — of the fraction of total polarization caused by galactic dust. Extrapolation from a more detailed map, released last month by the Planck team, suggests that the fraction is closer to 8–15%, Spergel explains.

With those updated numbers, he says, "there's no evidence for the detection of gravitational waves". But a final determination cannot be made until a more precise dust map, expected to be released by the Planck team in October, is available, he adds.

In the other analysis<sup>1</sup>, Seljak and Michael Mortonson, a cosmologist also at the University of California, Berkeley, re-examined BICEP2 data on how the polarization signal varies with the frequency of the microwaves it detects. The BICEP2 team had checked its results against data recorded at lower frequency by an older telescope, BICEP1. They found that the intensity of polarization did not change from one frequency to the other in the way expected if it were caused by dust, and concluded that the data favoured gravitational waves over dust by an 11-to-1 margin.

But Seljak and Mortonson say that the BICEP2 analysis did not exclude data on small spatial scales, or fractions of degrees of the sky.

That is a problem, Seljak says, because on these small scales, gravitational lensing — in which the path of light bends around massive objects — exactly mimics the twisting polarization pattern that gravitational waves imprint on larger spatial scales.

Accounting for lensing, "the primordial gravity-wave signal is preferred to dust with odds of less than two to one — in other words, not significant odds at all," says Seljak.

BICEP2 co-leader James Bock, a physicist at the California Institute of Technology in Pasadena, says that although his group's main paper "has been revised based on many referee comments and resubmitted" for publication, the evidence for gravitational waves "is certainly not being retracted". The BICEP2 results "are basically unchanged", he says.

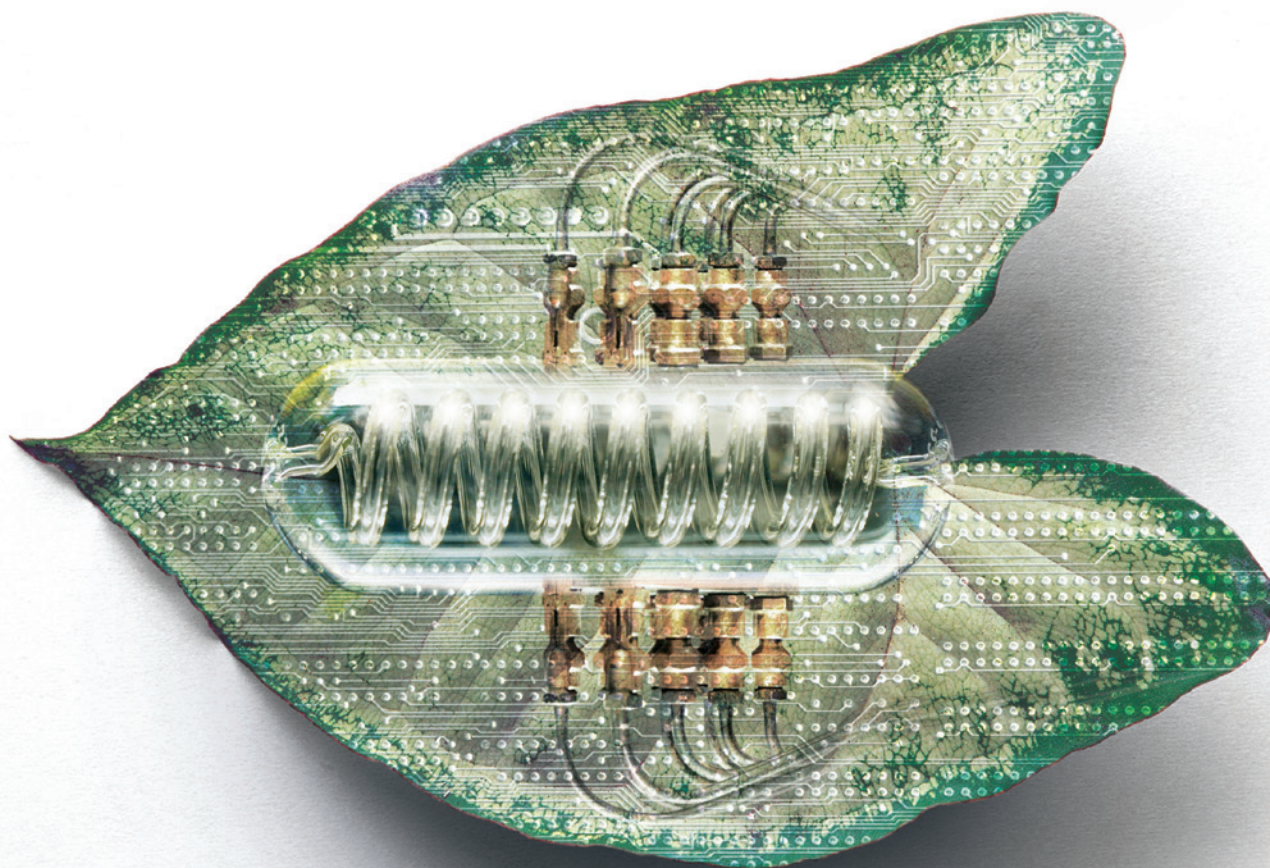
Further observations may yet see the cosmic ripples emerge from the dust. It is possible that forthcoming data from several observatories — including the Keck Array, a telescope at the South Pole built by the BICEP2 team — and the Planck team's full-sky map of CMB polarization will confirm that a signal is there, although perhaps not as strong as first suggested. ■

1. Mortonson, M. M. & Seljak, U. Preprint at <http://arxiv.org/abs/1405.5857> (2014).
2. Flauger, R., Hill, J. C. & Spergel, D. N. Preprint at <http://arxiv.org/abs/1405.7351> (2014).

## CORRECTIONS

The News story 'Jelly genome mystery' (*Nature* **509**, 411; 2014) wrongly stated that ctenophores might be the closest living relatives of the first animals. It should have said that ctenophores may represent the earliest lineage to split off from the common ancestor of all animals. The News story 'US Arctic research ship ready to cast off' (*Nature* **509**, 542–543; 2014) incorrectly stated that a deck on the vessel able to accommodate large equipment is at the bow — it is at the stern. The News Feature 'Complexity on the horizon' (*Nature* **509**, 552–553; 2014) inadvertently underplayed the role of Daniel Harlow in bringing computational complexity to fundamental physics — he worked with Patrick Hayden from the start of their project.





# SPRINGTIME FOR THE ARTIFICIAL LEAF

*Researchers make headway in turning photons into fuel.*

BY JESSICA MARSHALL

**O**n a bright spring morning in Pasadena, California, the air is rich with the smells of cut grass and flowers. Photosynthesis seems effortless here: the fronds and blooms that line the walkways of the California Institute of Technology (Caltech) bask in the sunlight, quietly using its energy to store sugars, stretch their leaves, deepen their roots and tend to their cellular processes.

Inside Caltech's Jorgensen Laboratory, however, more than 80 researchers are putting a lot of effort into doing the leaf's job using silicon, nickel, iron and any number of other materials that would be more at home inside a cell phone than a plant cell. Their gleaming new labs are the headquarters of the Joint Center for Artificial Photosynthesis (JCAP), a 190-person research programme funded by

ILLUSTRATION BY VIKTOR KOEN



the US Department of Energy (DOE) with US\$116 million over five years. The centre's goal is to use sunlight to make hydrogen and other fuels much more efficiently than real leaves ever made biomass.

The researchers are pursuing this goal with a certain urgency. Roughly 13% of greenhouse-gas emissions worldwide come from transportation, so phasing out polluting fuels is a key environmental target. One approach is to replace cars and light trucks with electric vehicles charged by solar cells or wind — but that cannot tackle the whole problem. Nathan Lewis, an inorganic chemist at Caltech and JCAP's scientific director, says that some 40% of current global transportation cannot be electrified. For example, barring a major breakthrough, there will never be a plug-in hybrid plane: no craft could hold enough batteries. Liquid fuels are unbeatable when it comes to convenience combined with compact energy storage.

That is why funding agencies around the world — and at least a few private companies — are putting unprecedented resources into making fuels using power from the Sun, which is not only carbon-free but effectively inexhaustible. JCAP stands out not only for its scale, but also for its ambition. It is one of five Energy Innovation Hubs created by the DOE beginning in 2010 to focus on specific problems using basic research, applied research and engineering. JCAP has promised to deliver a working prototype of an artificial leaf by the time its initial grant runs out in 2015.

Although the centre has taken some important steps in that direction — including one reported just last week<sup>1</sup> — it is still a long way from delivering on that promise. “This is a really, really difficult, challenging problem,” says electrochemist John Turner of the US National Renewable Energy Laboratory in Golden, Colorado. “The payback would be huge, but it's not as simple as everyone wanted it to be when we started playing in this area 40 years ago.”

Still, the surge of funding and attention has given many researchers reason to hope for long-term success. “If you could sustain this type of effort for the next ten years,” says Michael Wasielewski, a chemist at Northwestern University in Evanston, Illinois, “it's conceivable you could have a practical solution.”

## CATCHING RAYS

The concept of artificial photosynthesis goes back to 1912, but the push to achieve it did not start until 1972, when Japanese researchers outlined what a device would need to take in sunlight and use it to split water into oxygen and hydrogen fuel<sup>2</sup>. Progress was slow. In 1998, Turner reported<sup>3</sup> a complete system that showed a major advance — it stored 12% of the incoming solar energy as fuel, compared with 1% of energy stored as biomass in real leaves. But it cost more than 25 times too much to be competitive, and its performance dropped off after 20 hours of sunshine.

There are three things you want from an artificial leaf, says Lewis: “You want it to be efficient, cheap and robust. I can give you any two today, but not the third at the same time.”

JCAP's mission is to fix that problem — and in the process, to create a system that is much

**“THE PAYBACK  
WOULD BE HUGE,  
BUT IT'S NOT AS  
SIMPLE AS EVERYONE  
WANTED IT TO BE.”**

cheaper than just splitting water with electricity from a solar panel. At the heart of JCAP's artificial-leaf design are two electrodes immersed in an aqueous solution. Typically, each electrode is made of a semiconductor material chosen to capture light energy from a particular part of the solar spectrum, and coated with a catalyst that will help to generate hydrogen or oxygen at useful speeds (see ‘Splitting water’). Like many other artificial-photosynthesis devices, JCAP's system is divided by a membrane to keep the resulting gases apart and reduce the risk of an explosive reaction.

Once the water has been split, the hydrogen is harvested. It can be used as a fuel by itself — perhaps in hydrogen-powered cars such as those already making their way into showrooms in California — or be reacted with carbon monoxide to make liquid-hydrocarbon fuels.

Making any one of the artificial leaf's components work well is a challenge; combining all of them into a complete system is even harder. “This is exactly like building a plane,” says Lewis. “You've got to not just have an engine, you have to have a design with wings and the fuselage and the engine and the avionics — and the plane, in the end, has to fly.”

Much of the difficulty comes down to finding the right materials. Silicon, for instance, makes a good photocathode — the electrode that produces hydrogen gas — but is stable only when the solution around it is acidic. Unfortunately, the situation is reversed with photoanodes, which produce oxygen: the good ones are stable only when the solution is basic, not acidic. And the best catalyst for the oxygen-producing electrode, iridium, is both rare and expensive,

which makes it unsuitable for commercial-scale devices.

JCAP's High Throughput Experimentation lab is tackling the materials

problem with inkjet printers modified to churn out spots of alloys on glass plates for testing as catalysts and photoabsorbers. Together, the printers are able to produce up to one million spots of varying composition per day.

In one experiment<sup>4</sup> to find the best proportions of nickel, iron, cobalt and cerium oxides to generate oxygen from water, the team screened nearly 5,500 combinations for stability and function using a miniaturized chemical lab that glided over the glass plates tirelessly. The best-performing combination is not the most effective catalyst ever found for this reaction, but it is transparent, allowing light to pass through to the photoabsorber, and it has good chemical compatibility with that material.

One of the toughest challenges for artificial photosynthesis has always been getting a good material for the photoanode, says Carl Koval, an electrochemist and JCAP's director. “Those things were always horribly unstable, often not even stable for minutes.” Many researchers have focused their search on materials known to be cheap and stable — certain metal oxides, for example — and tried to make them into good light absorbers. Others feel that it is better to start with materials that are known to be efficient light harvesters, and to work at making them stable and cheap.

Just last week, a JCAP team reported<sup>1</sup> success with the latter approach. By putting a protective coating of titanium dioxide on high-performing photoabsorbers such as silicon, the researchers achieved big gains in stability. “That's basically the last piece of the puzzle to create the first-generation prototype,” says Koval, who predicts that JCAP will have an artificial leaf running in the next few months.

Publication of a preliminary system including the titanium dioxide coating is in the works, says Lewis. “That's going to be a double-digit-efficiency, stable system.” The threshold for commercial viability is thought to be in the 10–20% range. The photoabsorbers will not be cheap enough to bring to market, concedes Lewis, because their cores are made from expensive single-crystal silicon. But if subsequent research shows that cheaper fabrication methods work, the system could be cost-effective.

## SPECTRUM OF IDEAS

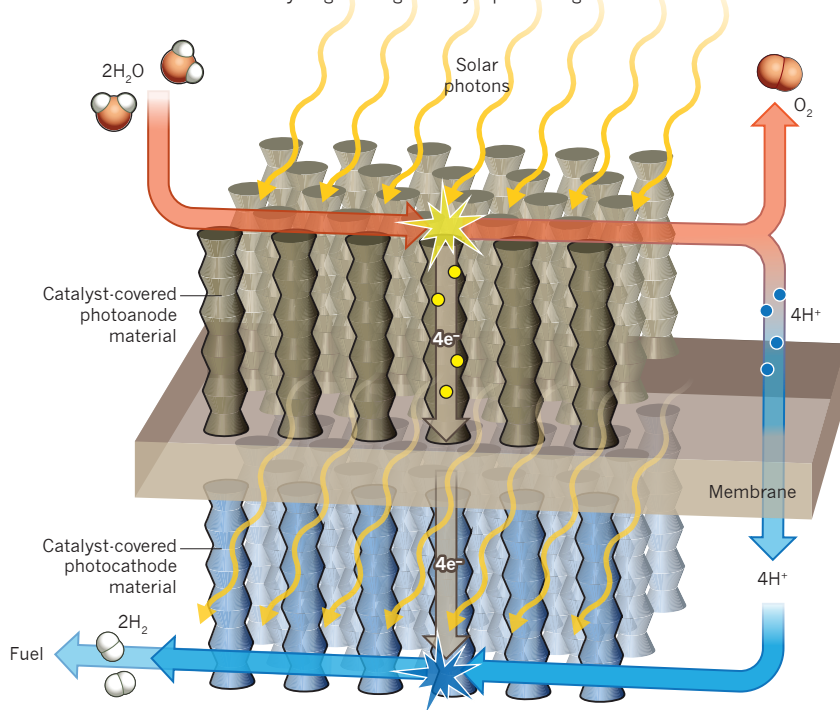
JCAP will soon complete its fourth year of operations. It got off to a slow start as new labs were built, but researchers both in and outside the centre praise its systematic focus on producing a practical system, and its progress so far. Even Turner, whose lab bid to become the solar-fuels hub but lost out to Lewis's team, is encouraged by JCAP's latest direction.

Still, says Koval, the centre has its critics. Some take issue with its emphasis on engineering and prototype development. But if JCAP were to focus on basic science, he says, it would not be “doing what the DOE created the hub for in the first place”. Other critics object to how JCAP concentrates on just one of several

➔ **NATURE.COM**  
Learn about another  
Energy Innovation  
Hub.  
[go.nature.com/2cdwk7](http://go.nature.com/2cdwk7)

## SPLITTING WATER

Artificial photosynthesis uses photons from sunlight to split water molecules into oxygen and hydrogen, which can be used to make fuel. Every two molecules of water yield one oxygen molecule ( $O_2$ ), as well as four pairs of protons ( $H^+$ ) and electrons ( $e^-$ ). The protons and electrons migrate across a membrane, where a photocathode recombines them into hydrogen using a catalyst plus sunlight.



possible ways to tackle artificial photosynthesis. “A lot of people would have been happier if the DOE had spread the funding around all these different ways of doing this,” says Koval. But that kind of dilution of effort would be risky in its own way, he argues: “Then you’d have progress on none of them.”

Besides, many of the alternative approaches are being pursued elsewhere. Up the coast in Santa Barbara, California, for example, a start-up firm called HyperSolar is testing a system in which coated nano- or micro-particles combining a photoabsorber and a catalyst are placed in a transparent, water-filled plastic bag. The bag will inflate as it is exposed to sunlight, and hydrogen and oxygen gas form inside. Such units could be deployed in sunny regions such as deserts. A 2009 DOE report<sup>5</sup> estimated that, if it uses cheap materials, this ‘baggie’ approach could produce hydrogen economically with 10% efficiency, stable for 10 years.

But the system is risky, says Turner, because it produces oxygen along with the hydrogen. “If you’re talking about 100 square miles of baggies in the desert with this explosive mixture,” he says, “one lightning bolt and you have a disaster.” HyperSolar researchers are exploring several ways to eliminate that danger. One is to use a system that separates the gases into two bags, says Syed Mubeen, a postdoc at the University of California, Santa Barbara, and lead scientist at the company. Another is to run the system using waste water rather than pure water, so that the oxygen reacts with organic

impurities and degrades them into valuable chemicals. This approach “completely removes oxygen out of the equation,” says Mubeen. As with JCAP’s stable photoanode, HyperSolar’s photoabsorber is protected by a coating.

### LIGHT INDUSTRY

Another entrant in the artificial-photosynthesis field is the Japan Technological Research Association of Artificial Photosynthetic Chemical Process (ARPChem), a consortium of universities and companies that has government funding comparable to JCAP’s grant — although over ten years rather than five — to develop a bag-based approach. Kazunari Domen, a chemist at the University of Tokyo and leader of ARPChem’s water-splitting group, says that one of the companies in the consortium has been working on a membrane to separate the hydrogen and oxygen products.

Other projects are making photoabsorbers from organic molecules, rather than semiconductors. Some are building molecular assemblies inspired directly by the photosynthetic apparatus of plants. And in the past few years, a class of materials called perovskites has drawn the attention of the solar-photovoltaic community for its high energy-conversion efficiency; some researchers think that the materials also have potential in artificial photosynthesis.

Daniel Nocera, a chemist at Harvard University in Cambridge, Massachusetts, launched Sun Catalytix to develop his work on a low-cost catalyst. But the company announced last year that it has put that

research on hold to pursue a less challenging product with prospects of turning a profit for investors sooner. The decision underscores the challenges of bringing a commercially viable artificial-photosynthesis system to market.

### BERKELEY BUBBLES

On a spring day in the arty-industrial district of Berkeley, California, researchers demonstrate a prototype system inside the temporary lab space that houses JCAP’s northern site. As a sunlamp shines on a CD-sized plastic box, fine streams of hydrogen bubbles rise between blue strips of catalyst-coated silicon and exit through tubes in the box’s top. This prototype system is not the team’s best: it won’t last and it is not very efficient. But it is still encouraging to see champagne-like bubbles triggered simply by light.

Then Karl Walczak, a postdoc in JCAP’s prototyping group, slides a second plastic box in front of the lamp. Inside is a small black square: a new titanium dioxide-coated photocathode. This second system immediately begins to generate bubbles much faster than the first. “This is where the field is going,” says Walczak.

JCAP researchers hope that such prototypes will ultimately lead to industrial hydrogen-production plants. They predict arrays of cells kilometres long, with a tower supplying water and pipes drawing the hydrogen to a storage tank. Some researchers propose that domestic units may also be part of the future, but Lewis warns that the small amount of sunlight that falls on a rooftop cannot make enough hydrogen to supply a family’s energy needs. Others say that the technology could be useful in areas of the developing world that lack an energy infrastructure, offering distributed fuel generation where it is needed.

In the meantime, researchers at JCAP and elsewhere are moving forward on all fronts. Devens Gust, a chemist at Arizona State University in Tempe, echoes a near-universal sentiment. “The bottom line,” he says, “is that nobody really knows yet what’s going to win out, what’s going to be practical.”

But whatever technology prevails, says Lewis, the logic behind artificial photosynthesis is inexorable. “The biggest energy source we have by far is the Sun,” he says. “The best way to store energy other than in the nucleus of an atom is in chemical fuels. It’s inevitable someone is going to take the biggest source and store it in the most dense way.” ■ [SEE EDITORIAL P.7](#)

**Jessica Marshall** is a freelance writer in St Paul, Minnesota.

1. Hu, S. *et al. Science* **344**, 1005–1009 (2014).
2. Fujishima, A. & Honda, K. *Nature* **238**, 37–38 (1972).
3. Khaselev, O. & Turner, J. A. *Science* **280**, 425–427 (1998).
4. Haber, J. A. *et al. Energy Environ. Sci.* **7**, 682–688 (2014).
5. James, B. D., Baum, G. N., Perez, J. & Baum, K. N. *Technoeconomic Analysis of Photoelectrochemical (PEC) Hydrogen Production* (Directed Technologies, 2009).



# The forgetting GENE

*For decades, most researchers ignored the leading genetic risk factor for Alzheimer's disease. That is set to change.*

BY LAURA SPINNEY

One day in 1991, neurologist Warren Strittmatter asked his boss to look at some bewildering data. Strittmatter was studying amyloid- $\beta$ , the main component of the molecular clumps found in the brains of people with Alzheimer's disease. He was hunting for amyloid-binding proteins in the fluid that buffers the brain and spinal cord, and had fished out one called apolipoprotein E (ApoE), which had no obvious connection with the disease.

Strittmatter's boss, geneticist Allen Roses of Duke University in Durham, North Carolina, immediately realized that his colleague had stumbled across something exciting. Two years earlier, the group had identified a genetic association between Alzheimer's and a region of chromosome 19. Roses knew that the gene encoding ApoE was also on chromosome 19. "It was like a lightning bolt," he says. "It changed my life."

In humans, there are three common

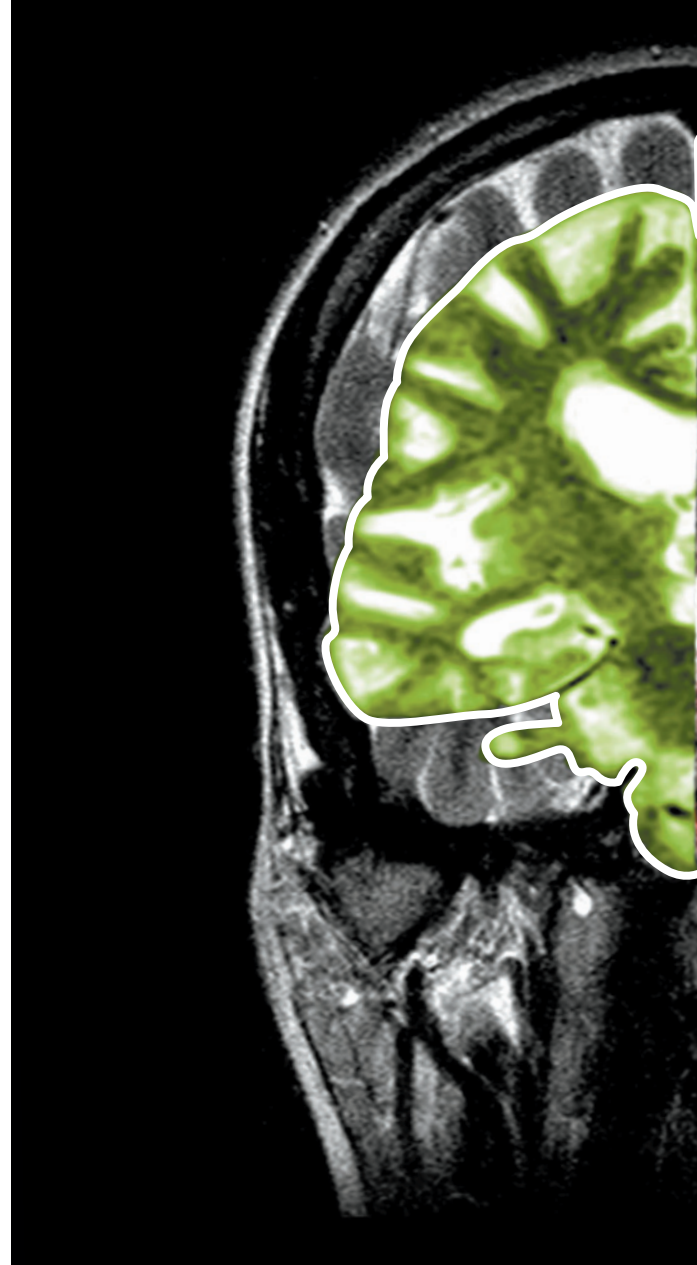
variants, or alleles, of the *APOE* gene, numbered 2, 3 and 4. The obvious step, Roses realized, was to find out whether individual *APOE* alleles influence the risk of developing Alzheimer's disease. The variants can be distinguished from one another using a technique called the polymerase chain reaction (PCR). But Roses had little experience with PCR, so he asked the postdocs in his team to test samples from people with the disease and healthy controls. The postdocs refused: they were busy hunting for genes underlying Alzheimer's, and *APOE* seemed an unlikely candidate. The feeling in the lab, recalls Roses, was that "the chief was off on one of his crazy ideas".

Roses then talked to his wife, Ann Saunders, a mouse geneticist who was skilled at PCR. She had just given birth to their daughter and was on maternity leave, so they struck a deal. "She did the experiments while I held the baby," he says. Within three weeks, they had

collected the data that would fuel a series of landmark papers showing that the *APOE4* allele is associated with a greatly increased risk of Alzheimer's disease<sup>1</sup>.

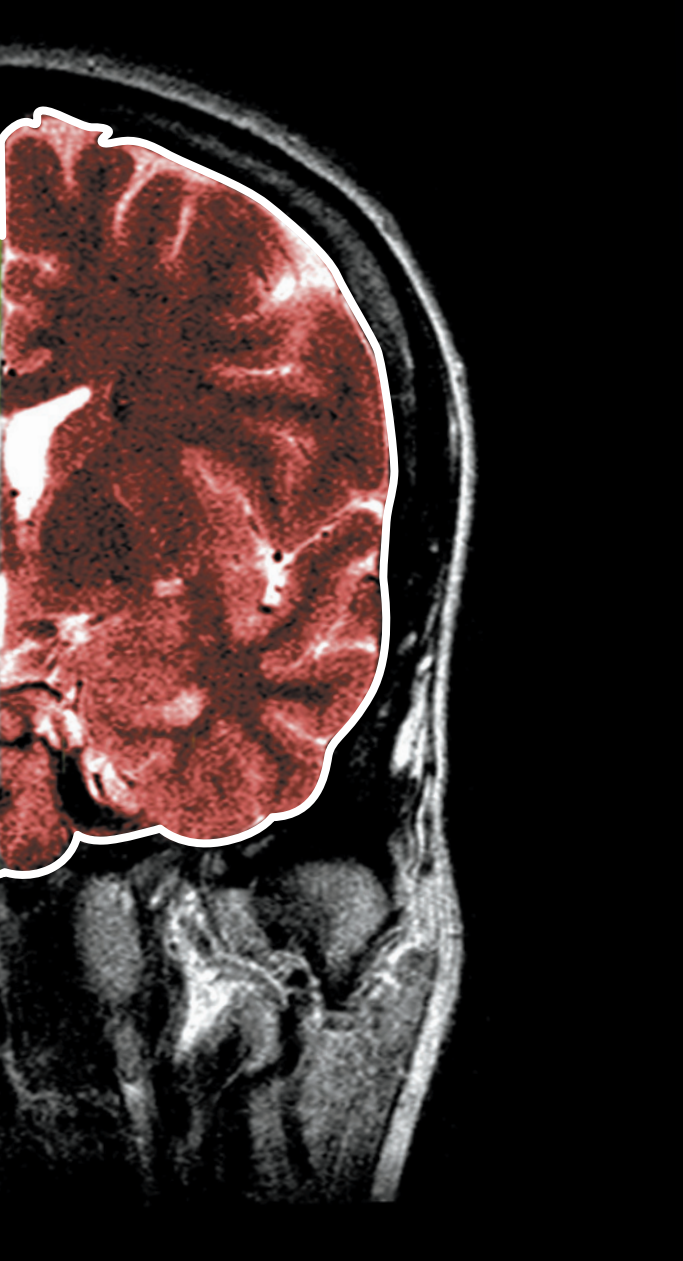
Twenty years on, *APOE4* remains the leading genetic risk factor for Alzheimer's, the most common form of dementia (see 'Risky inheritance'). Inheriting one copy of *APOE4* raises a person's risk of developing the disease fourfold. With two copies, the risk increases 12-fold. Yet Roses' data were largely criticized or ignored. Within a couple of years, interest in ApoE had dwindled as researchers flocked to study amyloid- $\beta$ . The handful of labs that continued to pursue ApoE did so in the face of indifference from funding agencies and the neuroscience community, and without the resources needed to validate experimental findings with larger studies.

Today, the function of the ApoE protein in the brain remains mostly unknown. This neglect of such a strong



MEDICAL BODY SCANS/JESSICA WILSON/SPL





Over time, Alzheimer's disease drastically shrinks the brain.

lead has puzzled some outside the Alzheimer's field. At a forum on brain diseases in Frankfurt, Germany, Thomas Bourgeron, an autism researcher at the Pasteur Institute in Paris, voiced his confusion. "If I had a risk factor like that, I'd be hot on its trail."

But interest in the lipoprotein is picking up, in part because attempts to target amyloid- $\beta$  have repeatedly disappointed in major clinical trials. Pharmaceutical companies are pulling back from amyloid-based approaches and some academics have begun to question the focus on the molecule. For the first time, researchers are developing drugs aimed at the ApoE4 protein and drawing attention from industry.

"The amyloid hypothesis became such a strong scientific orthodoxy that it began to be accepted on the basis of faith rather than evidence," says Zaven Khachaturian, president of the non-profit campaign Prevent Alzheimer's Disease 2020, and former coordinator

of Alzheimer's-related activities at the US National Institutes of Health. Until recently, he says, "no one has stepped back to ask the fundamental question of whether our basic premise about the disease is the correct one".

### STIFF COMPETITION

Opinions differ as to why Roses' finding was neglected, but many agree that bad timing played a part. In 1991, John Hardy and David Allsop had proposed the 'amyloid cascade hypothesis'. This posits that Alzheimer's disease results from the abnormal build-up of amyloid- $\beta$  clusters, or plaques, in the brain<sup>2</sup>. Others rallied around the idea and it has won most of the funding available to the field ever since.

But Roses did not subscribe to that theory. "Amyloid is one of many substances that builds up in plaques as a result of dying cells and atrophy in the brain," he says. "I never did think it was the cause." In saying so, he may have deterred others from investigating a possible ApoE-amyloid link, and inadvertently set up a competition between the two hypotheses for funding. He never got another grant to work on ApoE.

But there were also technical obstacles to ApoE research. The protein is found throughout the body, making it difficult to target the molecule specifically in the brain. And ApoE is bound to fat, so it tends to stick to other molecules in biochemical assays, says Menelas Pangalos, who leads research on small-molecule discovery at AstraZeneca in Macclesfield, UK, and has long had an interest in ApoE.

Working with such proteins requires an intimate understanding of lipid biochemistry. "If you want to study ApoE biology, you really need to devote a laboratory to understanding the techniques," says neurologist David Holtzman of Washington University in St. Louis, Missouri. Holtzman did just that, establishing a separate lab dedicated to developing techniques for handling lipoproteins in the central nervous system.

Amyloid was the easier target. Two decades of intensive pursuit have yielded a range of drugs that alter the metabolism of amyloid- $\beta$ , but these have yet to fulfil expectations. Of the six drugs that were in phase II or III clinical trials in 2012, half have since been dropped because of either safety concerns or lack of effectiveness. This comes against a backdrop of ageing populations, overstretched health-care systems and a dearth of medications for Alzheimer's disease. "The large number of major failed trials in Alzheimer's is

quite frightening," says Lennart Mucke, director of the Gladstone Institute of Neurological Disease at the University of California, San Francisco. "It has really scared off big pharma."

The three remaining drug candidates that target amyloid- $\beta$  are currently being tested in people with Alzheimer's, as well as in individuals who have a high risk of developing the disease but who have not yet developed symptoms. Imaging studies have shown that the brains of high-risk individuals look and behave differently from controls decades before the onset of Alzheimer's, and long before they start to accumulate amyloid- $\beta$  or lose grey matter<sup>3</sup>. The trials will examine whether the drugs prevent or delay the onset of the disease; they are due to wrap up over the next six years. There is a growing sense in the field — among academics and industry representatives alike — that these efforts are the last chance for the amyloid hypothesis. Amid these concerns, the spotlight has swung back to ApoE.

If the prevention trials fail, it will be up to academics to persuade companies back to the table with solid preclinical and early clinical data, says Mucke. He is optimistic that ApoE researchers may soon have that leverage. Despite the obstacles in this area, there is an emerging understanding of how ApoE4 increases risk, which Holtzman's and Mucke's groups have explored through transgenic mice that they have developed to express human forms of ApoE.

The molecule seems to contribute to Alzheimer's through two distinct pathways, one of which is amyloid-dependent. In both animals and humans, ApoE4 strongly promotes amyloid- $\beta$  deposition in the brain, compared with ApoE3, long considered the 'neutral' form when it comes to Alzheimer's risk. ApoE2, which is considered the protective form, decreases the build-up<sup>4</sup>. "These are compelling data," says Holtzman.

The other mechanism does not involve amyloid. When neurons are under stress, they make ApoE as part of a repair mechanism. The 'bad' ApoE4 form tends to be broken down into toxic fragments that damage the cell's energy factories — the mitochondria — and alter the cell skeleton.

The relative contribution of these two pathways to Alzheimer's risk is not known, says Holtzman, but he and others think that changing a harmful form of ApoE into a less damaging one might prove a promising therapeutic approach. At the Gladstone, cardiovascular scientist Robert Mahley, working with a team including neuroscientist

Yadong Huang, has identified small 'corrector' molecules that modify the structure of ApoE4 protein to one more like that of ApoE3, thereby reducing abnormal fragmentation<sup>5</sup>.

In cell culture, low concentrations of these corrector molecules can reduce mitochondrial impairment and neuronal dysfunction<sup>6</sup>. They are now being tested more rigorously in a range of animal models. If the molecules ultimately prove safe and effective in humans, Mucke foresees a day when doctors will prescribe them for people deemed at risk of Alzheimer's, just as statins are offered to those with high cholesterol and an elevated risk of cardiovascular disease.

### ABOVE AND BEYOND

Such drugs could also have implications beyond Alzheimer's. "The mitochondrial-impairment hypothesis provides a pretty logical and parsimonious explanation for why ApoE4 does bad things," says Mucke, "not only in the context of Alzheimer's, but maybe also in other diseases." There is evidence that it may be a risk factor in Parkinson's disease and epilepsy. It is also associated with an increased risk of a poor outcome after brain injury, and more rapid progression of untreated HIV infection. Fifteen biotechnology companies are already collaborating with the Gladstone to develop these and similar drugs.

Despite his inability to get grants, Roses never gave up on ApoE. But a few years after his group discovered the link between ApoE and Alzheimer's, he wearied of the constant battles for funding. He left academia and spent ten years in industry — where he continued to work on ApoE, among other things — before returning to Duke in 2008.

In 2009, his group described a stretch of non-coding DNA in a gene called *TOMM40* that sits next to *APOE* on chromosome 19. This stretch of DNA, known by the shorthand 523, varies in length. The length of 523 can determine the extent to which *TOMM40* and *APOE* are expressed<sup>7</sup>.

The discovery was important, Roses says, because the protein encoded by *TOMM40*, called Tom40, is crucial to healthy mitochondria. Tom40 forms a channel in the outer mitochondrial membrane that is used to import proteins. Without these proteins, mitochondria cannot divide as they should throughout a cell's life. "It's a big effect that's been known about for a decade," says Roses, "But it's not well-known in the Alzheimer's field."

Roses went on to suggest that 523 could be exploited to develop therapies

and improved tests for Alzheimer's risk. Most people will develop Alzheimer's if they live long enough, but only about 25% of people carry an *APOE4* allele. As a result, a prognostic test for *APOE4* will only ever be partially informative. But genotyping both *APOE* and *TOMM40* could provide information about a wider swathe of the population, Roses says. His group has found, for example, that *APOE3* — by far the most common *APOE* allele in humans — is usually inherited either with a short or a very long 523 tract. In those who inherit two *APOE3* alleles, age of onset differs depending on which combination of the two 523 variants they also inherit.

Other labs have found evidence supporting Roses' hypotheses, but some attempts to replicate his *TOMM40* findings have failed. In 2012, Hardy, now at University College London, and a colleague, geneticist Rita Guerreiro, wrote an editorial<sup>8</sup> in which they argued that *TOMM40* did not independently affect Alzheimer's risk.

Roses' faith in his hypothesis has not wavered: he believes he has a sound mechanistic explanation for his findings. And he says that the genome-wide studies that failed to reproduce his results lacked sufficient power to reveal the association between *TOMM40* and Alzheimer's disease. Khachaturian says a proper test of Roses' findings — using Roses' methods in a larger cohort of patients — has not yet been done.

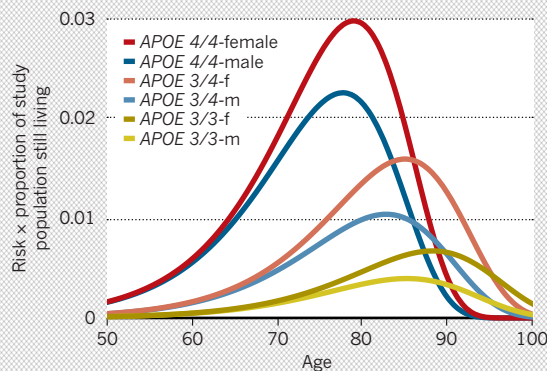
Roses hopes to soon be able to back up his findings with more clinical data and has launched a company called Zinfandel Pharmaceuticals in Durham, fuelled in part by his own funds. Along with the Japanese pharmaceutical company Takeda, based in Osaka, Zinfandel is currently funding a phase III trial, called TOMMORROW, that will put his ideas to the test. TOMMORROW is expected to run for about 5 years and will recruit close to 6,000 healthy, elderly individuals. It will evaluate a risk-assessment algorithm based on age, *APOE* and *TOMM40*.

The trial will also investigate whether a low dose of a drug called pioglitazone — already approved at much higher doses for certain patients with type 2 diabetes — can delay disease onset in those individuals deemed by the algorithm to be at high risk of Alzheimer's. Evidence from animal and small-scale human studies suggests that pioglitazone may prevent or reverse Alzheimer's-related pathology and symptoms<sup>9</sup>. Roses thinks it may do so by stimulating mitochondria to divide.

The ongoing trials could have major

### RISKY INHERITANCE

People who carry the gene variant *APOE4* tend to develop Alzheimer's at a younger age than those with two copies of *APOE3*.



**"No one has stepped back to ask whether our basic premise about the disease is the correct one."**

consequences even without yielding a cure: research has shown that an intervention that could delay the onset of Alzheimer's by just 2 years would result, 50 years later, in nearly 2 million fewer cases of the disease in the United States than projected otherwise<sup>10</sup>. And the results coming in over the next few years could force researchers to re-evaluate their understanding of dementia. It is time it was recognized for what it is, says Khachaturian: a failure of complex, interacting physiological systems. Looking at any one of these systems — even those involving ApoE4 — in isolation is unlikely to fully explain changes in behaviour. "The field is going to recognize the limitations of current approaches and step back," he says. "And if we're honest with ourselves, we'll start forging new directions." ■

**Laura Spinney** is a freelance writer in Lausanne, Switzerland.

1. Strittmatter, W. J. et al. *Proc. Natl Acad. Sci. USA* **90**, 1977–1981 (1993).
2. Hardy, J. & Allsop, D. *Trends Pharmacol. Sci.* **12**, 383–388 (1991).
3. Filippini, N. et al. *Proc. Natl Acad. Sci. USA* **106**, 7209–7214 (2009).
4. Kim, J., Basak, J. M. & Holtzman, D. M. *Neuron* **63**, 287–303 (2009).
5. Mahley, R. W. & Huang Y. J. *Med. Chem.* **55**, 8997–9008 (2012).
6. Chen, H. K. et al. *J. Biol. Chem.* **287**, 5253–5266 (2012).
7. Linnertz, C. et al. *Alzheimers Dement.* <http://dx.doi.org/10.1016/j.jalz.2013.08.280> (2014).
8. Guerreiro, R. J. & Hardy, J. *Arch. Neurol.* **69**, 1243–1244 (2012).
9. Sato, T. et al. *Neurobiol. Aging* **32**, 1626–1633 (2011).
10. Brookmeyer, R., Gray, S. & Kawas, C. *Am. J. Pub. Health* **88**, 1337–1342 (1998).



# COMMENT

**HISTORY** William Astbury, another overlooked figure in the DNA-structure story **p.32**



**MEDICINE** A chronicle of shifting attitudes to US childhood obesity **p.34**

**DATA** Citizen science from previous centuries is ripe for mining **p.35**

**OBITUARY** Gerald Guralnik, Higgs-boson physicist, remembered **p.36**

IMAGINECHINA/CORBIS



In Lanzhou, China, 700 mountains are being levelled to create more than 250 square kilometres of flat land.

## Accelerate research on land creation

China's campaign to bulldoze mountains to build cities needs expertise to avert geoengineering problems, warn **Peiyue Li, Hui Qian and Jianhua Wu**.

China's cities are expanding rapidly as its economy grows<sup>1</sup>. Yet land for development is in short supply, particularly in mountainous areas, where about one-fifth of the population lives<sup>2</sup>. In the past decade, local governments have begun removing the tops of mountains to fill in valleys to create land on which to build.

This is reminiscent of an ancient Chinese fable, 'The Foolish Old Man who Removed the Mountains'. In the tale, a 90-year-old

man convinces his disbelieving neighbour that he can dig away, stone by stone, two mountains that block the way from his house. Because he succeeds (albeit with the help of deities) the fable is often cited — including by Mao Zedong — to illustrate the power of perseverance. But in our view, China should heed the story's title: earth-moving on this scale without scientific support is folly.

In cities such as Chongqing, Shiyang, Yichang, Lanzhou and Yan'an, tens of square

kilometres of land have been created<sup>3</sup> (see 'Moving mountains'). One of the largest projects, which started in April 2012 in Yan'an in the Shaanxi province, will double the city's current area by creating 78.5 square kilometres of flat ground<sup>3</sup>. Local officials are optimistic that the sale or leasing of the new land will generate billions of renminbi, while easing pressures on agriculturally or culturally valuable land elsewhere.

But the consequences of these ►

► unprecedented programmes have not been thought through — environmentally, technically or economically. There has been too little modelling of the costs and benefits of land creation. Inexperience and technical problems delay projects and add costs, and the environmental impacts are not being thoroughly considered. Yan'an, for example, is the largest project ever attempted on loess, thick million-year-old deposits of wind-blown silt. Such soft soils can subside when wet, causing structural collapse<sup>4</sup>.

Overcoming these problems will require the sharing of experiences and technical strategies nationally and internationally. Collaboration between universities, research institutes, private companies, construction teams and governments is needed to ensure that the best research informs the engineering, so that the vast sums of public money involved are invested wisely. Without cooperation, the land-creation campaigns could drain local and national economies and irrevocably damage watercourses and ecosystems.

### TOO LITTLE, TOO LATE

Land creation by cutting off hilltops and moving massive quantities of dirt is like performing major surgery on Earth's crust<sup>5</sup>. Mountaintop moving has been done before in strip mining, especially in the eastern United States<sup>6,7</sup>. But it has never been carried out on this scale. In China, dozens of hills 100–150 metres in height are being flattened over hundreds of kilometres. Such infill has never been used for urban construction. There are no guidelines for creating land in the complex geological and hydrogeological conditions that are typical of mountainous zones.

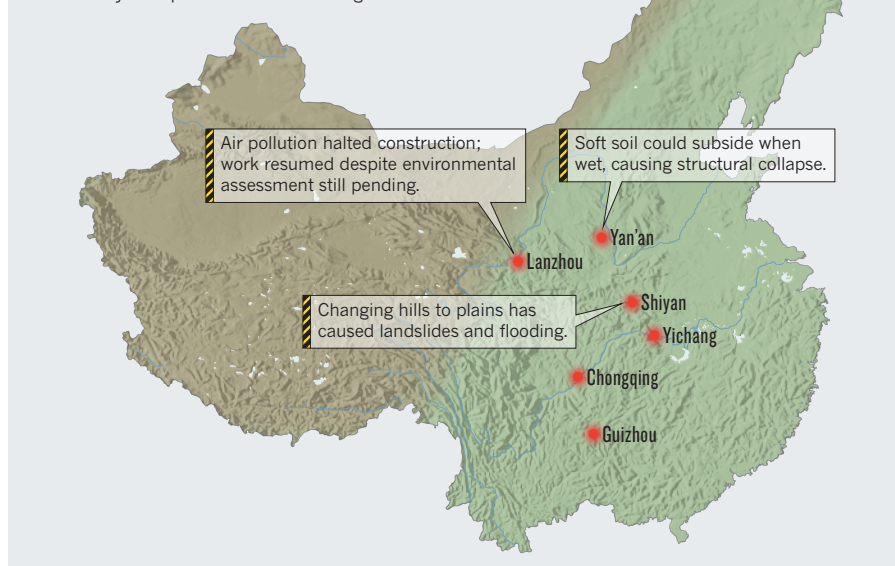
Land-creation projects are already causing air and water pollution, soil erosion and geological hazards such as subsidence<sup>3,5</sup>. They destroy forests and farmlands and endanger wild animals and plants. The city of Shiyang, for example, lies near the headwaters of the South–North Water Transfer Project — a major initiative to divert water from rivers through canals in southern China to Beijing and the north. There, the changing of hills to plains has caused landslides and flooding, and altered watercourses. This causes greater soil erosion, increasing the sediment content of local water sources (see [go.nature.com/klasxh](http://go.nature.com/klasxh); in Chinese).

In Yan'an, the air is often brown with dust owing to construction teams working on windy days without dampening the soil. Forests and plants on hills and in gullies are stripped ahead of the demolition and filling.

**“Government funds are not allocated according to scientific significance or need.”**

### MOVING MOUNTAINS

Tens of square kilometres of land have been created by levelling hills in several cities in mainland China. Work began in 2012 to make nearly 80 square kilometres of flat ground in Yan'an.



This increases the risk of soil erosion and groundwater loss, because farmlands and forests block wind and retain moisture and soil grains. During the earth-moving project in the city of Lanzhou<sup>8</sup>, soil erosion is expected to increase by 10% and concentrations of dust particles in the air by 49%.

Many land-creation projects in China ignore environmental regulations, because local governments tend to prioritize making money over protecting nature. When air pollution associated with the Lanzhou project became visible in April 2013, work was halted pending an environmental assessment. Four weeks later, construction resumed because of mounting costs to the local government and contractors. The assessment has yet to be completed.

The assumption that land creation always brings economic benefits needs to be challenged through cost modelling. For example, the 100-billion-renminbi (US\$16-billion) total cost over ten years of the Yan'an plan could take decades to recoup. The infill area cannot be built on for at least another decade, when its ground base becomes stable. This long payback period will dissuade some investors from buying the land, thus increasing economic risk.

### INADEQUATE INSIGHT

Research into land creation is insufficient. A lack of expertise, collaboration and funding means that scientists have not been able to conquer technical problems or provide tested, efficient or timely support for the engineering projects. Basic research into the hydraulic parameters of soils and aquifers was stalled by a year-long delay in distributing government research funding for 2013.

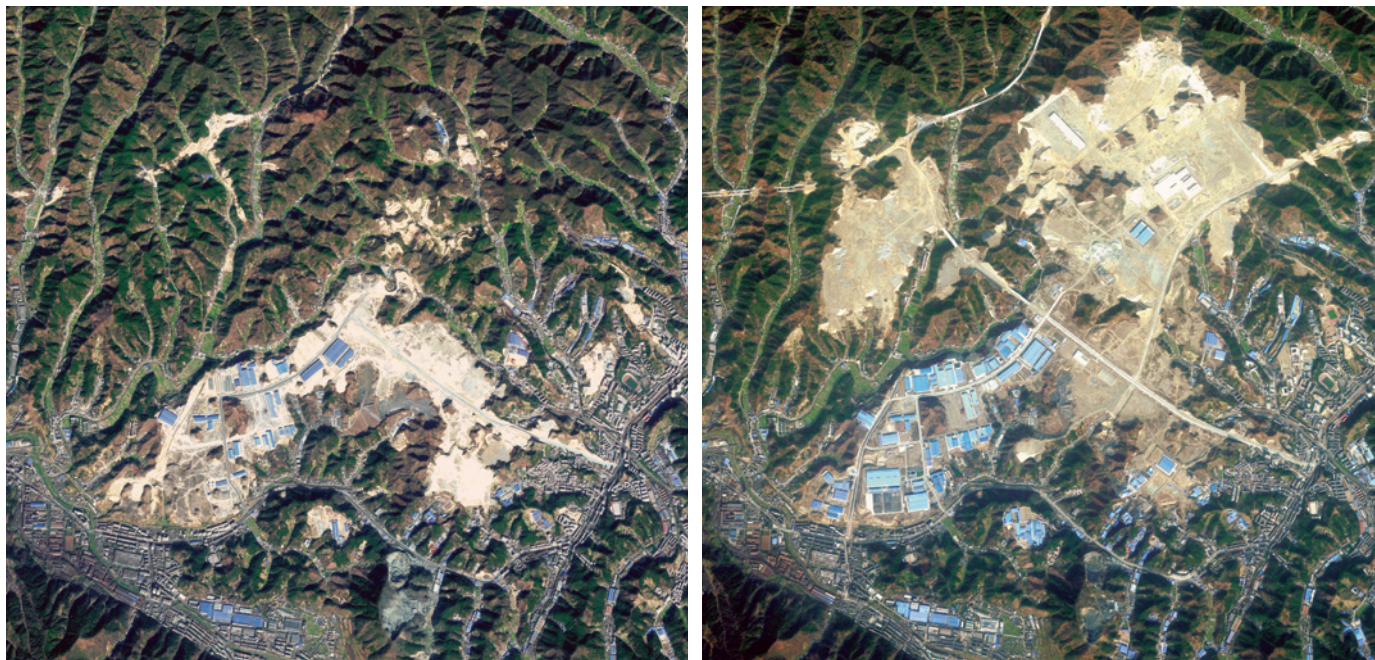
Research is local and narrow. In Yan'an, the Chinese government has granted 30 million renminbi to several local institutes and universities to solve some technical problems. The research focuses on the monitoring and modelling of ground deformation, subsidence, hydrogeological changes and protection of the surrounding slopes. Economics and environmental risks are not included.

The number and breadth of research teams is inadequate, and the limited experience and skills of local scientists are a problem. Researchers in Yan'an, for example, have expertise in hydrogeology and engineering geology but are less experienced in soil-water research, which is important for the stability of the ground base. Institutes have struggled to collaborate. Each manages its own sub-project and seldom shares research data with others. Teams must repeat work to obtain data that another institute already has, wasting time and money.

No international organizations or scientists who have rich expertise and practical experience are participating. Those who have participated in mountaintop removal in the United States should be invited to take part in Chinese projects.

Research funding comes exclusively from the government. Industries, institutes and universities carry out the research, yet few governmental departments are involved, limiting their influence. In Yan'an, only the Shaanxi Science and Technology Department and the Chinese Ministry of Science and Technology have participated or provided funding. Other local government departments involved in areas such as housing or environmental





Satellite images of western Shiyao, China, in 2010 (left) and 2012 (right) after several peaks have been flattened.

protection are not involved.

Government funds are not allocated according to scientific significance or need. Rather, they are directed on discretion of the local institute in charge of the project. For instance, water-soil interaction is the key factor affecting the success of the Yan'an project but the drive to understand it attracts only 12% of the research budget. And environmental and ecology-related research receives no funding in Yan'an, despite being very much in the public interest.

Research also lags behind the engineering work, and so has little impact. In Yan'an, the research started three months after excavations began. Lab tests that could have established the exact moisture content needed to harden loess foundations were unavailable to guide the project. Preliminary results are now available, but for only a few soils. An earlier start could have delivered more.

### BOOST RESEARCH

We call on the Chinese government to accelerate research efforts urgently. With good guidance, land creation could proceed with fewer risks and more benefits.

China's best researchers from Beijing, Shanghai and Wuhan should join regional collaborations. And a national network, similar to the National Basic Research Programs of China, should be established to coordinate this effort, to ensure quality and avoid duplication.

Scientists from the US Geological Survey, the US Environmental Protection Agency, the International Association of Hydrogeologists and universities in Canada, the United States and Europe should also be brought in. Local construction industries

could be involved and provide funds and equipment.

Local and national governments should allocate funds according to scientific importance, guided by a pool of national experts including hydrogeologists, environmentalists, economists and government officials. Extra funding must be found for environmental and ecological risk assessment and for economic studies of land creation. An independent management organization should be established and supervised by local government to manage and share the research data.

Environmentalists, ecologists, hydrogeologists, engineers and decision-makers must collaborate to find solutions to environmental problems — including the disappearances of small creeks, endangering of wild animals and birds and altered groundwater flows.

As a minimum the following recommendations should be heeded: during earth movement, soils should be watered to raise the moisture content to avoid dust pollution; ground cover and trees on hills and in gullies that are destroyed before infilling should be replanted as soon as possible to prevent soil erosion; and farmers who lose farmland should be compensated.

Economists need to assess the costs and benefits of land-creation projects and suggest ways to reduce the risks before these projects begin. Where high economic risks and low profitability are predicted, projects should not proceed even if they are

technically feasible, at least for now. Where there is no profit, governments should be dissuaded from going ahead.

Consulting environmental scientists and economists in advance will prevent or minimize extra costs arising from delays or troubleshooting. Pausing the Yan'an excavations would cost 500,000 renminbi a day in payments to idle construction workers, for instance<sup>9</sup>.

As in the ancient fable, China is moving mountains. Until we know more about the consequences, we urge governments to seek scientific advice and proceed with caution. ■

**Peiyue Li** is assistant professor of hydrogeology and environmental science in the School of Environmental Science and Engineering at Chang'an University, China. **Hui Qian** and **Jianhua Wu** are in the School of Environmental Science and Engineering at Chang'an University, China. e-mail: lipy2@163.com

1. Allen, F., Qian, J. & Qian, M. J. *Finan. Econ.* **77**, 57–116 (2005).
2. Li, M. in *The Population of Modern China* (eds Poston, D. L. Jr & Yaukey, D.) 83–112 (Springer, 1992).
3. Qian, W. *China Newsweek* **2**, 26–29 (2013; in Chinese).
4. Sun, J. Z. *Loessology* Vol. 1 (Hong Kong Archaeological Society, 2005).
5. Song, Q. *China Construction* **20**, 21–23 (2013; in Chinese).
6. Burns, S. S. *Bringing Down the Mountains: The Impact of Mountaintop Removal Surface Coal Mining on Southern West Virginia Communities* (West Virginia Univ. Press, 2007).
7. Crotty, M. G. *Villanova Environ. Law J.* **13**, 287–313 (2002).
8. Zhang, M. Q., Ji, S. N. & Sun, N. X. *J. Arid Land Res. Environ.* **23**, 77–81 (2009; in Chinese).
9. Zhao, F. *China Bus. J.* **A13**, 1–5 (2013; in Chinese).





William Astbury used X-ray crystallography to unravel the changeable structures of proteins.

# MOLECULAR BIOLOGY

## Of DNA and broken dreams

Georgina Ferry weighs up a life of William Astbury, who had a forgotten role in pinning down the double helix.

History is written by the victors, in science as in war. The story of the race to uncover the structure of DNA was thrust into the public imagination by James Watson's highly partial 1968 account, *The Double Helix*. Ever since, historians have been filling in the gaps. Most notably, successive authors have corrected

Watson's unkind portrait of Rosalind Franklin, whose X-ray-diffraction photograph of DNA ('Photo 51') was crucial to the model developed by Watson with Francis Crick.

Now science historian Kersten Hall has disinterred another figure from the footnotes of the DNA story, and brought him into the limelight. William Astbury — known to

all as Bill — was part of the group of X-ray crystallographers who learned their craft at the feet of William Henry Bragg, the father of the field, at London's Royal Institution in the 1920s. Astbury, along with fellow student J. D. Bernal, was fired with the new technique's possibilities for studying the key molecules of life, especially the proteins that are fundamental to all living tissue.

Bernal left the Royal Institution to found a lab at the University of Cambridge, UK; Astbury was banished (as he saw it at the time) to the University of Leeds in northern England. There he applied X-ray analysis to the study of natural fibres such as wool and hair: Leeds was an important centre of the textile industry. The wool protein keratin, Astbury complained on his appointment, was "lifeless and uninteresting". Later, he would tell public audiences that wool had given him "a glimpse of the loom on which the web of life is woven".

In the early 1930s, Astbury obtained X-ray-diffraction photographs of stretched and unstretched wool fibres, which suggested that the protein chain could switch between a compact and an extended form. Apart from shedding light on the elasticity of wool, his discovery showed that the properties of long-chain proteins depended on their three-dimensional shape, and that changes in shape could have functional consequences.

Hall makes a good case for Astbury's contribution to protein-structure analysis, but the book's subtitle signals that the author is really interested in DNA. In 1938, Astbury and his PhD student Florence Bell — a protégée of Bragg's son Lawrence, and now largely forgotten — published the first X-ray photographs of DNA fibres. They were blurred and hard to analyse, because the Leeds team did not know that two forms of DNA were mixed in their sample. But they contributed to Watson and Crick's thinking, and encouraged Maurice Wilkins — and subsequently Franklin — at King's College London to undertake X-ray studies of the molecule.

In 1951, Astbury's "lab boy" turned doctoral student, Elwyn Beighton, produced X-ray photos of DNA that were almost identical to Franklin's Photo 51, taken the following year. Astbury never published Beighton's pictures or presented them in public, nor did he seem to recognize that the distinctive X-shaped diffraction pattern showed that DNA was a helix.

Hall is clearly frustrated that Astbury missed the chance to win a Nobel Prize and



**The Man in the Monkeynut Coat: William Astbury and the Forgotten Road to the Double-Helix**  
KERSTEN T. HALL  
Oxford University Press: 2014.

LEEDS UNIV. LIB.



change the face of biology. He dwells at length on what might account for this failure. Astbury had become depressed and embittered as he saw others in London and Cambridge win funding that he had failed to secure. He had conceded the DNA problem to Wilkins at King's. He may not have kept up with the theory of X-ray diffraction. Finally, Hall suggests, Astbury probably saw DNA's apparently fixed structure as uninteresting compared to the conformational changes in proteins; at the time, his main interest (and the subject of Beighton's thesis) was bacterial flagella.

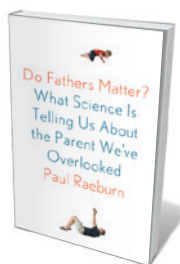
Hall indulges in some counterfactual speculation. What if Astbury had shown Beighton's pictures to the US chemist — and later Nobel laureate — Linus Pauling, who stayed with Astbury in Leeds in 1952? Pauling might have seen their significance and realized Watson and Crick's worst fears by beating them to the structure. But it didn't happen, and Watson's 1968 book credits Astbury only with the “one half-good photograph” taken by Bell in 1938.

Hall tells his story with style and pace. But I am unconvinced by his central contention, that Astbury was a “titan” and the founder of molecular biology. Astbury was one of many who independently promoted a molecular approach to understanding living things. Success in science depends as much on personality as on intellect, and here Hall leaves us largely in the dark. Clearly Astbury was a larger-than-life figure, a good communicator and entertaining in company (if you liked “off-colour jokes”), yet not favoured by the scientific establishment. Crystallography pioneer Dorothy Hodgkin (who taught chemistry to Astbury's daughter Maureen at the University of Oxford, UK — a connection not mentioned by Hall) once described him affectionately as “very bad and very amusing”.

And the monkeynut coat? Astbury worked with the now defunct UK company Imperial Chemical Industries to develop fibres made from a protein derived from peanuts, and proudly wore a coat made from ‘Ardil’. It proved to be no cheaper and considerably less hard-wearing than wool, and never took off. Hall uses this story to launch a final chapter exploring Astbury's prescient reflections on the manipulation of biological materials for utilitarian ends. He did not live to see the age of biotechnology, but he was certainly one of its prophets. ■

**Georgina Ferry** is a science writer based in Oxford, UK, and author of biographies of Dorothy Hodgkin and Max Perutz.  
e-mail: [mgf@georginaferry.com](mailto:mgf@georginaferry.com)

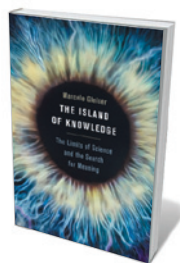
## Books in brief



### Do Fathers Matter?: What Science Is Telling Us About the Parent We've Overlooked

Paul Raeburn FARRAR, STRAUS & GIROUX (2014)

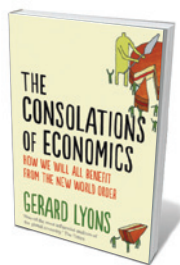
When science journalist Paul Raeburn first assessed the available research on fathers, he found the pickings decidedly slim. Scientists are now redressing the balance, and Raeburn has rounded up key findings in developmental psychology, evolutionary biology, genetics and neuroscience. Prepare for a bracing walk through the myriad ways fathers matter, from the “genomic battle of the sexes” that can lead to differing syndromes in offspring, to a father's ‘destabilizing’ — and hence stimulating and educative — play with his children.



### The Island of Knowledge: The Limits of Science and the Search for Meaning

Marcelo Gleiser BASIC BOOKS (2014)

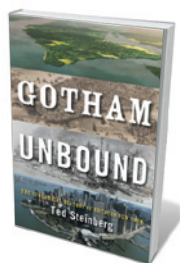
Among the bright knowns of science are countless unknowns, untouched by the most exquisitely calibrated, powerful instruments yet invented. Physicist Marcelo Gleiser sets out to explore the extent to which we can make sense of reality. He maps humanity's “island of knowledge” from pre-Socratic cosmology to quantum mechanics. And he marks how mysteries proliferate with each finding — a perpetual blur in our cosmic view that, like mathematician Kurt Gödel's work on incompleteness, spurs scientific creativity.



### The Consolations of Economics: How We Will All Benefit from the New World Order

Gerard Lyons FABER AND FABER (2014)

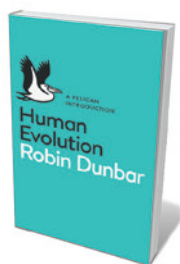
The economic conflagration of 2008 inspired plenty of bravura ideas on how to fireproof the future. Economist Gerard Lyons, who saw it all coming, offers a gloom-free reading of causes and correctives, and predicts global growth led by multiple economies. Drawing on key case studies and long experience with China, he examines world economic drivers, ‘soft’ and ‘hard’ power (a country's influence exerted through cultural appeal, or by military or economic means), and the economic goals and strategies that foster global stability.



### Gotham Unbound: The Ecological History of Greater New York

Ted Steinberg SIMON & SCHUSTER (2014)

How did Mannahatta, a wild expanse of hills and mudflats on North America's Atlantic coast, become the high-density, concrete-coated city of New York? Ted Steinberg's environmental history traces the stages from Henry Hudson's 1609 discovery of a “drowned estuary” to the ravages of Hurricane Sandy in 2012. Centuries of extreme land-use changes — reclaiming underwater terrain, levelling hills, draining lakes and crafting ersatz terra firma from landfill — have created an ecologically scarred city vulnerable to further storms. Other coastal megacities should take heed.



### Human Evolution: A Pelican Introduction

Robin Dunbar PELICAN BOOKS (2014)

The British Academy's ‘Lucy to Language’ project has spawned major sociobiological findings, most recently aired in *Thinking Big* by Robin Dunbar, Clive Gamble and John Gowlett (see *Nature* **509**, 284–285; 2014). Dunbar draws on that research — in particular the social-brain hypothesis and time-budget models — for this solid primer on human evolution under the relaunched Pelican imprint. On the journey from the australopithecines to the Neolithic and what followed, Dunbar is an accomplished guide. *Barbara Kiser*



Standards for children's weight began to be established at the turn of the twentieth century.

## HEALTH

# The medicalization of fat

A history of the shifts in US society's view of the larger child grips **David Katz**.

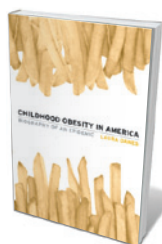
Between the moment I began to read Laura Dawes's *Childhood Obesity in America* and my completion of this review, two high-profile studies came out on US childhood obesity. The first suggested a slight decline in obesity rates among 2–5 year olds, spawning headlines claiming that overall childhood-obesity rates had plummeted (C. L. Ogden *et al.* *J. Am. Med. Assoc.* **311**, 806–814; 2014). The second reported that severe obesity in children aged 2–19 was rising (A. C. Skinner and J. A. Skelton *J. Am. Med. Assoc. Pediatr.* <http://doi.org/svw>; 2014).

Such wild gyrations say something about our emotionally charged relationship with this topic. However intense the preoccupation with childhood obesity, it is not an exclusively modern phenomenon. What differ are the medical, cultural and moral lenses through which we view it, and its salience in modern epidemiology. Dawes acknowledges that much of the prevailing attention to childhood obesity is “a lament of modernity”, and she has 100 years of carefully collated evidence to prove it. Through anecdotes and an overview of social trends, she provides the historical context for our fixation, revealing shifting cultural perceptions, medical preoccupations, scientific advances and economic forces linked to the phenomenon. By

clarifying where we have been, Dawes aims to guide us forward.

Adiposity has at times been admired. In Renaissance Europe, it was synonymous with wealth; in the Baroque period it was an attribute of feminine beauty. And it was valued in the United States around the start of the twentieth century, as Dawes shows. Charley Bilcher, for instance, was a Pennsylvanian boy whose bulk was a source of pride to his parents — and to the nation when, in 1896, he won the title of “fattest boy in the world”.

As Dawes shows, when the extra-large child was rare, she or (mostly) he was likely to be acclaimed. With nearly one-fifth of US children and adolescents now obese, the tide of opinion has turned. A tendency to appreciate bulk may have origins in the evolutionary struggle to ensure that children had enough to eat. Around the time of Bilcher's win, the girth of US-born versus immigrant children was a national preoccupation. “Better Babies”



**Childhood Obesity in America: Biography of an Epidemic**  
LAURA DAWES  
Harvard University Press: 2014.

contests became popular, fuelled in part by the “more abundant comforts” proffered by US industry. By showing how recently such a stance prevailed, Dawes shines a light on an important source of cultural ambivalence about the ‘war’ on obesity. Maybe we are yet to win because we are fighting with half a heart.

Dawes shows how much time and effort in the early twentieth century was devoted to generating measures of obesity that we take for granted today. A focus on length and weight standards was established with physician Luther Emmet Holt's 1894 *The Care and Feeding of Children*, and measures such as the Ponderal index — precursors to the body mass index — emerged. The more reliably obesity could be measured, the more it mattered: measuring informs norms; norms are used for comparison; and comparisons illuminate outliers. Dawes traces the emergence of paediatrics as a medical speciality in the 1870s, advances in epidemiology such as the capture of population-level data to establish reference standards for height and weight, and a growing respect for objective measures.

To me, the most fascinating portion is the chronological sequence of ideas about the causes and cures of childhood obesity, because they reveal how often we have gone back to the future. In *Diet and Health* (1918), the first best-selling US diet book, physician Lulu Hunt Peters aired the idea of the calorie as a factor in obesity. The glandular theory of obesity was a prominent paradigm into the 1940s: a 1937 case series showed that eight obese children had been successfully treated with pituitary-gland extracts and thyroid tablets. Today, debate still simmers about the importance of the endocrine system in obesity, with some arguing that hormonal imbalance explains the condition so well that calories cease to matter. Dawes's chronology also shows how there has been a focus on one cause or remedy to the exclusion of others. When ‘glands’ were to blame, psychology was largely ignored. When psychoanalysts held sway, endocrinologists were pushed aside. In Dawes's final chapters, she hints at a potential solution through a synthesis of such ideas.

Dawes is neither a clinician nor a public-health practitioner; she is a historian of medicine. But perhaps health professionals' efforts to contain, control, prevent and reverse childhood obesity have been constrained by a failure to see and consider the full scope of the threat and the best defences — that is, by the focus on one idea at a time. Dawes replaces such parochial perspectives with a window 100 years wide. May it help show us the way. ■

**David Katz** is the founding director of Yale University Prevention Research Center in Derby, Connecticut, and editor-in-chief of *Childhood Obesity*. His latest book is *Disease Proof*.  
e-mail: [david.katz@yale.edu](mailto:david.katz@yale.edu)



# Correspondence

## Design animal studies better

Steve Perrin argues that animal models of human disease should be better characterized to avoid false-positive results (*Nature* **507**, 423–425; 2014). Poor design, execution and reporting of studies are pervasive contributors as well.

For example, randomization and blinding in animal studies is pitifully infrequent (see, for instance, J. P. Ioannidis *et al.* *Lancet* **383**, 166–175; 2014). Randomly allocating animals to experimental and control groups makes the experimental groups as similar as possible in other respects. Blinding promotes comparable handling and measurement by experimenters, and publicly preregistering animal-study protocols and the outcomes to be measured would identify and reduce bias in reporting results.

Creating largely homogeneous experiments aids reproducibility and boosts statistical power, but has a cost of generalizability: the few drugs that have translated successfully from animals are effective across a broad range of circumstances (see, for example, E. S. Sena *et al.* *J. Cereb. Blood Flow Metab.* **30**, 1905–1913; 2010).

Funders and ethics committees need to ensure that study designs include these features, and journals should make them criteria for publication if they have not done so already.

**Malcolm R. Macleod\*** *University of Edinburgh, UK.*

*malcolm.macleod@ed.ac.uk*

*\*On behalf of 6 co-authors (see [go.nature.com/js4uzj](http://go.nature.com/js4uzj) for full list).*

## Investment suggests RNAi is on the up

If pharmaceutical investment is a measure of the prospects for therapeutics that use RNA-interference (RNAi) technology for gene silencing (*Nature* **508**, 443; 2014), on balance, things are looking up.

Yes, Novartis shuttered its RNAi effort and Merck sold its RNAi programme to Alnylam Pharmaceuticals. These were the right decisions: Novartis and Merck are both accomplished companies, but their decade-long efforts in RNAi failed to yield any clinical programmes.

What you did not mention was Genzyme's investment in Alnylam in early 2014 — one of the largest and most significant strategic transactions in recent biotech history. Genzyme, a Sanofi company, purchased a 12% stake in our company in return for broad product rights in the field of rare diseases in regions outside North America and western Europe.

Companies such as Genzyme do well to partner with technology innovators such as Alnylam for access to their products, not to the technology platform. This is how drug companies ultimately had success with recombinant DNA and monoclonal antibodies. The same story could be emerging with RNAi.

**John Maraganore** *Alnylam Pharmaceuticals, Cambridge, Massachusetts, USA.*  
*jmaraganore@alnylam.com*

## Mine centuries-old citizen science

Historical data on biodiversity would be valuable for investigating the long-term impact of human activities. Contrary to popular belief, such data have been widely collected for several hundred years through initiatives that would today be described as 'citizen science'.

For example, in the late sixteenth century, the Spanish government distributed questionnaires known as *relaciones topográficas* to each village, with local inhabitants providing a compilation of natural-history knowledge. The 637 questionnaires that survive

include information on some 190 species of wild animals and plants, gleaned from more than 4,300 individual records.

Also, the 16-volume geographical dictionary edited by the statistician and politician Pascual Madoz in the mid-nineteenth century contains information on most Spanish population centres, rivers and geographical landmarks. Madoz's enterprise involved the collaboration of more than 1,000 citizens and provided several thousand records of wild plants and animals, as yet unused.

Similar historical data sets also exist from China, most European countries and their former colonies. These can be biased towards socio-economically important species, but this bias can help to minimize misidentifications, for instance.

We call for an international, multidisciplinary effort (including historians, linguistic experts, geographers and biologists) to locate, compile and contextualize these invaluable historical data for incorporation into global biodiversity databases.

**Miguel Clavero, Eloy Revilla** *Doñana Biological Station — CSIC, Seville, Spain.*  
*miguelclavero@ebd.csic.es*

## Moon mapped by an artist's impression

Thomas Harriot mapped the Moon before Galileo (J. Rampling *Nature* **508**, 39–40; 2014). Unlike Galileo, he did not interpret what he saw as craters, mountains and valleys (see S. Pumfrey *Notes Rec. R. Soc.* **63**, 163–168; 2009).

Galileo's experience in the visual arts gave him a theoretical background in mathematical perspective (see, for example, S. Y. Edgerton *The Heritage of Giotto's Geometry* 233–239, Cornell Univ. Press, 1991; and S. E. Booth and A. Van Helden *Sci. Context* **13**, 463–486; 2000).

This included using shadows to determine the volume and location of an object. Observing the Moon's surface with his *perspicillum*, later named telescope, Galileo's training as a draftsman led him to interpret variations in light on the surface as shadows produced by the Sun on a complex topography.

Harriot did not grasp the significance of these dark areas. He interpreted them simply as surface markings that he used to 'map' the Moon. Art and science have sometimes benefited each other in ways that we fail to appreciate.

**José Rafael Martínez Enríquez** *National Autonomous University of Mexico, Mexico City, Mexico.*  
*enriquez@unam.mx*

## Experimentation needs theory, too

John Skoyles emphasizes the importance of practical experimental work for the developing scientist (*Nature* **508**, 319; 2014). But theory is crucial too — for interpreting the results and for advancing research.

In 1928, British physicist Paul Dirac came up with his equation, which predicted the existence of the positron. The particle was discovered independently four years later by US physicist Carl Anderson at the California Institute of Technology in Pasadena. This is a good example of "theory guides; experiment decides" (the motto of analytical chemist Izaak Maurits Kolthoff).

Or as the astrophysicist Arthur Stanley Eddington declared: "It is also a good rule not to put overmuch confidence in the observational results ... until they are confirmed by theory."

Perhaps the last word should come from Soviet physicist Peter Leonidovich Kapitza. He said: "Theory is a good thing, but a good experiment is forever."

**Min-Liang Wong** *National Chung-Hsing University, Taiwan.*  
*mlwong@dragon.nchu.edu.tw*

# Gerald Guralnik

## (1936–2014)

Physicist who helped to conceive the Higgs boson.

Gerald Guralnik is best known for his work on the Higgs mechanism, which confers mass on particles. Along with particle physicist Carl Hagen and myself, he wrote one of three key papers published in 1964 that each took different routes to solving a conundrum of how particles interact.

Guralnik died of a heart attack on 26 April, aged 77, shortly after finishing a lecture at Brown University in Providence, Rhode Island, where he had been a faculty member since 1967. He was born in Cedar Falls, Iowa, where his parents ran an accountancy firm. Guralnik attended the Massachusetts Institute of Technology (MIT) in Cambridge. He then went on to Harvard University, also in Cambridge, to study for a PhD under Walter Gilbert, who was working on the theory of elementary particles before embarking on his Nobel-prizewinning experiments for DNA sequencing.

In 1964, Guralnik completed his doctorate in quantum field theory, which models subatomic particles as physical fields that extend throughout space. That same year, he won a US National Science Foundation fellowship to a theoretical physics research group, founded by Gilbert's former adviser Abdus Salam, at Imperial College London.

In the years after the Second World War, the success of quantum electrodynamics, which describes the electromagnetic force that drives interactions between matter and light, had spurred work on similar theories of other interactions. A key goal was a unified theory of the electromagnetic force and the weak force, which is observed in radioactive  $\beta$ -decay and the Sun's core. The big problem was that to explain the short range of the weak forces, the bosons transmitting them had to have large masses. By contrast, the electromagnetic force is carried over vast distances by the massless photon.

When I met Guralnik in 1964 at Imperial, I was interested to learn that he, too, was concerned with the problem of boson mass. Hagen, his friend and former MIT classmate, also came to Imperial, and the three of us worked together to find a solution. We were constant lunch companions. Guralnik enjoyed the work but often complained about the food — especially Scotch eggs (hard-boiled eggs baked into a grainy coating of sausage meat and breadcrumbs).

Just as our paper was in its final draft, we discovered two earlier papers (one by

François Englert and Robert Brout at the Université Libre de Bruxelles in Belgium, and the other by Peter Higgs at the University of Edinburgh, UK) that addressed the problem of giving mass to bosons from



different perspectives. All three papers showed that the mechanism could successfully use spontaneous symmetry breaking, in which the underlying physics is symmetric but its realization is not. For example, a bowl of water is rotationally symmetric, until it freezes and the ice crystals break the symmetry. A particular contribution of our paper was to show how spontaneous symmetry breaking could occur consistently within the formal structure of the theory. The three papers reached essentially the same conclusions, and were later selected by *Physical Review Letters* (the journal in which they were published) as among the most significant of 1964.

The publications attracted few citations in their first three years. Guralnik toured Europe and gave talks to sceptical audiences. Werner Heisenberg, known for formulating the eponymous uncertainty principle, told him that he was talking nonsense. When Guralnik won a postdoc position at the University of Rochester in New York, the head of the group urged him to switch to a different field (later admitting that the advice was misguided).

Our proposal, now generally known as the Higgs mechanism, is the cornerstone of the well-established unified electroweak theory,

which explains both the electromagnetic and weak interactions. Steven Weinberg and Salam married this mechanism with a preliminary unified model proposed earlier by Sheldon Glashow, thus resolving its inconsistencies. For this work, the three of them shared the 1979 Nobel Prize in Physics.

The validity of the Higgs mechanism was finally demonstrated in 2012, with the experimental discovery of the Higgs boson, which last year led to another Nobel Prize, for Englert and Higgs, the surviving authors of the first two papers putting forth the mechanism.

The 1979 prizewinners carefully cited all three of the 1964 papers on an equal footing, but others have not always done so. Guralnik came to feel that our early paper was often unfairly neglected. He gave talks and wrote papers pointing out our distinctive contribution, of which he was justifiably proud, and in which he was unquestionably the prime mover.

In the meantime, Guralnik had gone on to contribute much to the development of computational approaches to quantum field theory. From 1985 to 1987, as a staff member at Los Alamos National Laboratory in New Mexico, he used such approaches to make important contributions to quantum chromodynamics, the theory of strong interactions, which explains that quarks interact through a 'colour force'. Later, he worked on how chaos theory could be applied to particle physics. He also made significant contributions to string theory, some in collaboration with his son, Zachary Guralnik.

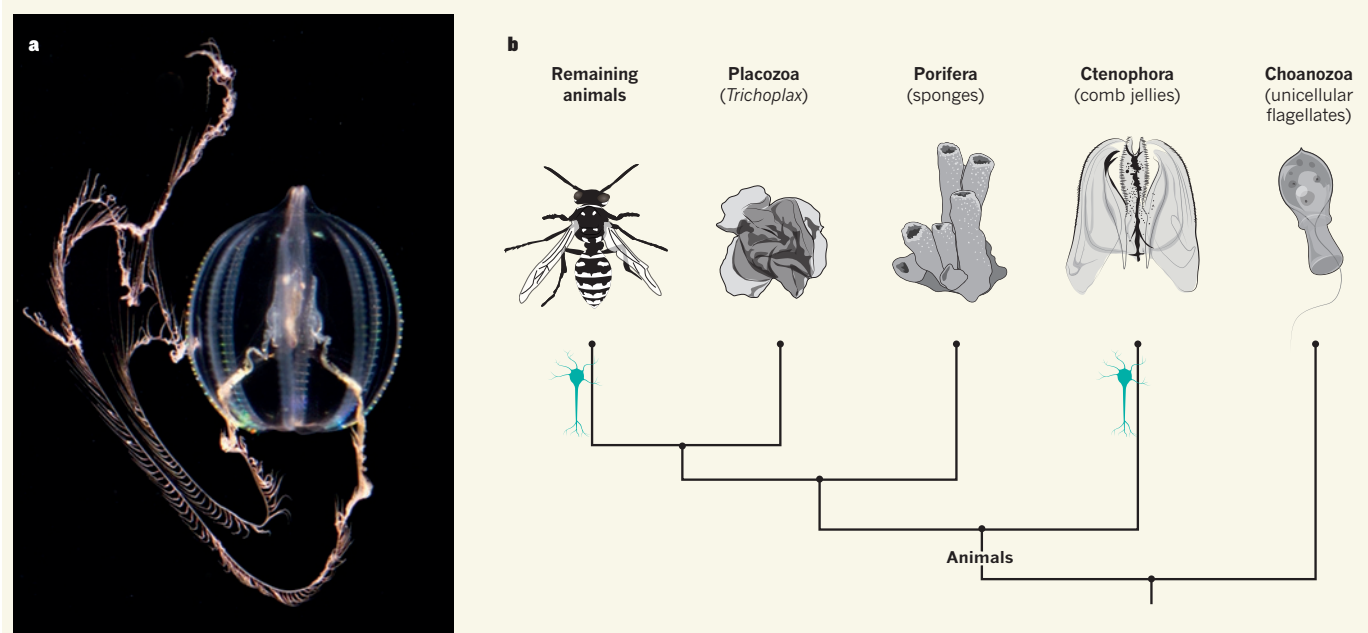
Gerry was always good company and had a great sense of humour. As well as being passionate about physics, he was enthusiastic about many things: fast cars, photography and state-of-the-art computers. Gerry was a member of Brown's Ersatz Brain Project, which aims to design a brain-like computer. He could be combative when others failed to live up to his exacting standards, but he was warm and generous.

Gerry once wrote that being able to work as a theoretical physicist was a privilege; his colleagues felt the privilege was theirs. ■

*Tom Kibble is emeritus professor of theoretical physics at Imperial College London. He worked there with Gerald Guralnik in the 1960s.  
e-mail: kibble@ic.ac.uk*

MIKE COHEA/BROWN UNIV.





A. LEONID L. MOROZ & MATHEW CITARELLA

**Figure 1 | Dual origins of nerves.** **a**, The comb jelly *Pleurobrachia bachei*. **b**, An evolutionary tree depicting the genetic relationships between the early animal lineages. Moroz *et al.*<sup>1</sup> suggest that nerve cells (indicated by the neuron symbol) have evolved independently in the lineage that includes humans and the one that leads to comb jellies. Nerve cells are absent in all other animal groups.

## EVOLUTIONARY BIOLOGY

# Excitation over jelly nerves

Analysis of the draft genome of a comb jelly and of gene-transcription profiles from ten other ctenophores hints at an independent evolutionary origin for the nervous systems of these organisms. [SEE ARTICLE P.109](#)

ANDREAS HEJNOL

The ability of animals to respond rapidly and appropriately to changes in their environment is due to the presence of a nervous system consisting of up to billions of nerve cells. In this issue, Moroz *et al.*<sup>1</sup> (page 109) present the genome of the comb jelly *Pleurobrachia bachei*, otherwise known as the sea gooseberry (Fig. 1a). Following a detailed examination of the developmental genes, structural genes and signalling molecules that are necessary for the set-up and function of nervous systems in other animals, the authors come to a radical conclusion: that the nervous system of comb jellies might have evolved independently of that of other animals.

Only two groups of animals do without a nervous system: sponges, which are simple animals attached to the sea bottom that do not show complex behaviours, and the placozoans, animals comprised of two flat sheets of cells that creep along the ocean floor absorbing nutrients. The simplicity of sponges and placozoans has led generations of zoologists to

conclude that they are ancient animal groups, and may look very like the first multicellular animals that emerged on the planet more than 500 million years ago.

Over the past decade, however, extensive comparisons of protein and DNA sequences have led to surprising rearrangements at the base of the animal tree of life. In fact, it seems that previous assumptions about the origin of multicellular animals may be wrong, and that a group of gelatinous creatures, the ctenophores, collectively referred to as comb jellies, could be the first group to have branched off from the animal tree of life<sup>2</sup>. Further evidence for this reappraisal came from a whole-genome analysis of another comb jelly, *Mnemiopsis leidyi*, which indicated that ctenophores lack many molecular characteristics that are essential for all other animals<sup>3</sup>. For example, they do not have microRNAs, important controllers of gene expression in other animals<sup>4</sup>, or *Hox* genes, which encode transcription factors that are essential for patterning the body axis<sup>5</sup>.

So what are these mysterious animals that

have turned the table on textbook ideas? Comb jellies are fabulous marine predators that propel themselves through the water column by means of blocks of cilia — the shimmering combs that give them their name. They catch their prey using innervated tentacles seamed with sticky cells called colloblasts and swallow it through their mouth, which opens into a sac-like gut. They have a nerve net with regional specializations, such as a sensory organ located at one pole of the body that is used for light reception and gravity sensing<sup>6</sup>. Notably, sponges and placozoans lack all of these features, so the proposal that the jellies evolved first seems odd.

However, Moroz *et al.* agree with the idea of ctenophores being one of the two first animal lineages (Fig. 1b). Their analysis, which alongside the *Pleurobrachia* genome includes gene-transcription profiles from ten other ctenophores, finds that many building blocks essential for nervous-system development in other animals are also missing from comb jellies. Furthermore, some molecules that are neuron-specific in other animals are

present in the sea-gooseberry genome but, surprisingly, are expressed not in nerve cells but in other tissues. And of the 10 major common neurotransmitter molecules used by animals, *Pleurobrachia* employs only two (glutamate and its antagonist GABA) in the operation of its nervous system. On the basis of these characteristics, it would seem that an animal with such a small number of traditionally neural proteins would have a simple nerve net, as opposed to a central nervous system, and would not show any complex behaviour. On the contrary, however, comb jellies perform complex actions such as predation and horizontal diurnal migrations in the water column, so they must use different molecules in their nervous system.

The phylogenetic position of comb jellies at the base of the animal tree of life and the findings of Moroz and co-workers suggest a fascinating scenario — that comb jellies evolved a nervous system that is unrelated to that of other animals. Heretical hypotheses such as this strike a blow against the anthropocentric view that complex animals emerged gradually along one lineage only, culminating in humans, and that complex organ systems did not evolve twice. But such views do not reflect how evolution really works. Evolution does not follow a chain of events in which one lineage progresses continuously towards complexity while other branches stagnate<sup>7</sup>. Instead, it is an ongoing process in all lineages. When the animal tree branched more than 500 million years ago, one lineage gave rise to ctenophores and the other to all remaining animals alive today, and it seems that the two lineages independently evolved a rapid internal communication system.

However, the last word has not yet been said on this issue, because the branching sequence of the earliest animal groups is still hotly debated. Some researchers have expressed doubt that ctenophores are at the base, and claim that the lack of many genes in comb jellies can be explained by massive gene loss that mimics a simple genome<sup>8</sup>.

Regardless of where ctenophores finally end up on the tree, the development and evolution of the complex nervous system of these creatures will be an enigma for some time. If it turns out that comb jellies are not at the base of the tree and that animal neurons indeed originated only once, someone must figure out why the molecular biology underlying the comb-jelly nervous system is so different from that of other animals. ■

**Andreas Hejnol** is at the Sars International Centre for Marine Molecular Biology, University of Bergen, 5008 Bergen, Norway. e-mail: andreas.hejnol@sars.uib.no

1. Moroz, L. L. *et al.* *Nature* **510**, 109–114 (2014).
2. Dunn, C. W. *et al.* *Nature* **452**, 745–749 (2008).
3. Ryan, J. F. *et al.* *Science* **342**, 1242592 (2013).

4. Maxwell, E. K., Ryan, J. F., Schnitzler, C. E., Browne, W. E. & Baxevanis, A. D. *BMC Genom.* **13**, 714 (2012).
5. Ryan, J. F. *et al.* *EvoDevo* **1**, 9 (2010).
6. Jager, M. *et al.* *J. Exp. Zool. B* **316B**, 171–187 (2011).

7. Maxmen, A. *Nautilus* No. 9; <http://nautilus.us/issue/9/time/evolution-youre-drunk> (2014).
8. Dohrmann, M. & Wörheide, G. *Integr. Comp. Biol.* **53**, 503–511 (2013).

This article was published online on 21 May 2014.

## CLIMATE SCIENCE

# How Antarctic ice retreats

**New records of iceberg-rafted debris from the Scotia Sea reveal episodic retreat of the Antarctic Ice Sheet since the peak of the last glacial period, in step with changes in climate and global sea level. [SEE LETTER P.134](#)**

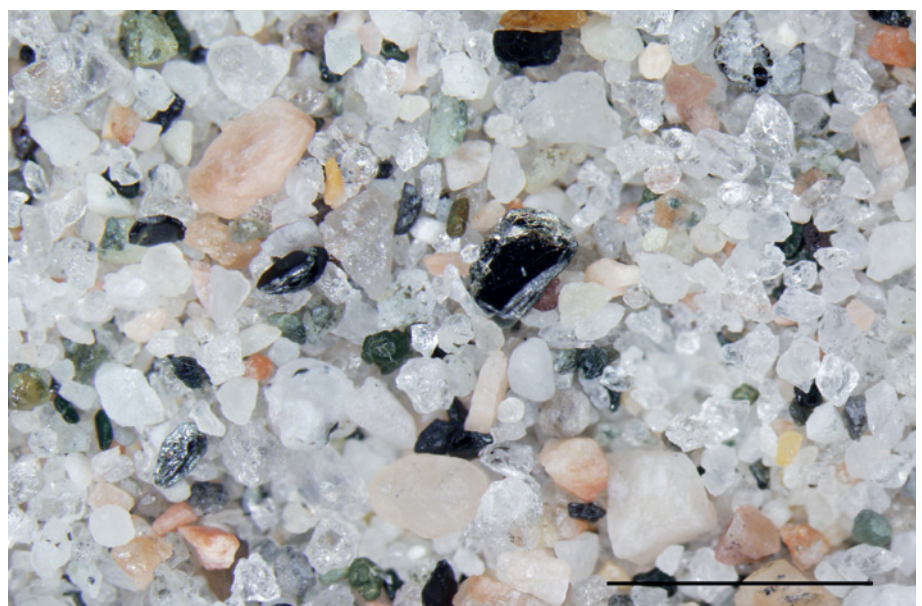
TREVOR WILLIAMS

About 19,000 years ago, the ice sheets that covered large areas of North America and Eurasia began to melt. The ice on Antarctica also melted, but to a lesser extent — most of it still exists today. Understanding past ice-sheet instability and melting is important for predicting future ice behaviour in a warming world. On page 134 of this issue, Weber *et al.*<sup>1</sup> present new well-dated records of iceberg-rafted debris from two marine sediment cores from the Scotia Sea that reveal at least eight episodes of ice loss from Antarctica between 20,000 and 9,000 years ago. These records allow examination of interactions between temperature, ice melt and the water masses of the Southern Ocean, which are

central to the carbon cycle and to climate change between glacial and interglacial periods.

Since the Last Glacial Maximum (LGM) — the peak of the most recent glacial period — which occurred between about 26,000 and 19,000 years ago<sup>2</sup>, melting ice sheets have raised global sea level by about 130 metres (ref. 3). But there are significant uncertainties in the timing and amount of ice lost from Antarctica. For example, estimates of sea-level rise resulting from Antarctic ice melt have ranged from about 8 m to 30 m, with the most recent estimates around the lower end of this range<sup>4</sup>.

Currently, Antarctica loses ice by two main processes: melting of the underside of floating ice shelves and calving of icebergs. The icebergs melt slowly as they are carried westwards along the coast of Antarctica, and icebergs



**Figure 1 | Iceberg-rafted debris from marine sediment offshore of Porpoise Bay, East Antarctica.** Weber *et al.*<sup>1</sup> describe similar iceberg-rafted debris from two marine sediment cores from the southern Scotia Sea that document several episodes of ice loss from Antarctica between 20,000 and 9,000 years ago. Scale bar, 2 mm.



that reach the Weddell Sea are blocked by the Antarctic Peninsula and turn northwards, where they melt more rapidly in the warmer waters of the Scotia Sea. Icebergs themselves are ephemeral, but they carry mineral grains and rock fragments scoured from Antarctic bedrock. As icebergs melt, this iceberg-rafted debris (IBRD; Fig. 1) falls to the seabed and is steadily buried in marine sediments to form a record of iceberg activity.

Some icebergs are more debris-rich than others: large tabular icebergs calved from floating ice shelves have already lost much of their debris-rich bases to melting, whereas icebergs that calve close to the line between grounded and floating ice tend to retain their debris, and are probably more common during ice-sheet retreat than during times of ice-sheet stability. Nevertheless, the main interpretive link is sound — more IBRD is a sign of more icebergs and greater ice loss from the Antarctic Ice Sheet. The two IBRD records reported by Weber and colleagues are similar, even though the sediment-core sites are separated by 2° of latitude and are subject to different local oceanographic conditions. This similarity increases confidence that the IBRD records represent an iceberg signal.

Until now, the timing of ice retreat has been constrained by radiocarbon dating of marine sediments and by dating of land surfaces that were uncovered as the ice sheets thinned<sup>5,6</sup>. Here, the eight episodes of iceberg discharge were dated by matching the record of wind-blown dust in the same sediment cores to wind-blown dust in an already-dated Antarctic ice core. This approach results in a continuously dated record, which provides a significant advance in knowledge of when the Antarctic Ice Sheet retreated over the time since the LGM. Weber *et al.* find that the first of the Antarctic ice discharges took place 20,000–19,000 years ago and was followed by a series of larger episodes between 17,000 to 9,000 years ago.

The largest iceberg release lasted from 14,800 to 14,400 years ago and overlapped, within dating uncertainty, with a period of sea-level rise known as meltwater pulse 1A (MWP-1A), which occurred 14,650 to 14,310 years ago<sup>7</sup>. During this period, sea levels rose by about 14–18 m at the astonishing rate of 4 m or more per century. Weber and colleagues' iceberg-discharge data clearly show a contribution to MWP-1A from Antarctica, but how much meltwater does this represent? Recent work<sup>4</sup> puts the total budget for sea-level rise from Antarctic ice melt since the LGM at about 9 m, and this melt budget has to be shared among the eight iceberg-discharge events, including MWP-1A. By this reckoning, the Antarctic contribution to MWP-1A is relatively minor compared with the contribution from Northern Hemisphere ice, which must provide the balance of the roughly 14–18-m rise in sea level. This result contrasts with modelling of differences in amplitude of the

MWP-1A sea-level rise at different locations (caused by the gravitational and rotational effects of removing the ice mass), which estimates that half or more of MWP-1A comes from Antarctica<sup>7</sup>. This mismatch has yet to be fully resolved.

Questions remain about which areas of the Antarctic Ice Sheet became unstable and produced these iceberg-discharge events. Did major ice-drainage sectors retreat simultaneously, perhaps in response to external triggering such as an initial sea-level rise from the north or southward migration of relatively warm Circumpolar Deep Water, as modelled by the authors? Or did different sectors retreat independently, as individual thresholds for instability were crossed in each sector? Principal sources of icebergs were probably the nearby Antarctic Peninsula and Weddell Sea embayment, where ice streams drain about a quarter of Antarctic ice by area. Icebergs are also likely to have travelled from other ice outlets around East Antarctica. The provenance of the IBRD, and the icebergs that carried it, can be found by matching the geochemical fingerprint (such as characteristic argon-isotope ages) of individual mineral grains in the IBRD to the corresponding geochemical fingerprint of the different source areas<sup>8</sup> — a topic for future study.

The episodic iceberg discharges described by Weber and colleagues shed light on the question of how the Antarctic ice sheets melt. Will there be similar iceberg releases in the future? Ice streams in the Amundsen Sea sector of the West Antarctic Ice Sheet are already

in the early stages of retreat<sup>9,10</sup>. During the last interglacial period, about 125,000 years ago, sea levels reached 6–9 m higher than today<sup>11,12</sup>, much of this attributable to an Antarctic meltwater source, at global temperatures only 1–2 °C warmer than those of today. The planet is on course for a temperature rise exceeding this value, so we can expect similar ice-sheet instability and retreat to that described by Weber *et al.* in the future. ■

**Trevor Williams** is at the Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA.  
e-mail: trevor@ldeo.columbia.edu

1. Weber, M. E. *et al.* *Nature* **510**, 134–138 (2014).
2. Clark, P. U. *et al.* *Science* **325**, 710–714 (2009).
3. Austermann, J., Mitrovica, J. X., Latychev, K. & Milne, G. A. *Nature Geosci.* **6**, 553–557 (2013).
4. Whitehouse, P. L., Bentley, M. J. & Le Brocq, A. M. *Quat. Sci. Rev.* **32**, 1–24 (2012).
5. Heroy, D. C. & Anderson, J. B. *Quat. Sci. Rev.* **26**, 3286–3297 (2007).
6. Hillenbrand, C.-D. *et al.* *Quat. Sci. Rev.* <http://dx.doi.org/10.1016/j.quascirev.2013.07.020> (2013).
7. Deschamps, P. *et al.* *Nature* **483**, 559–564 (2012).
8. Pierce, E. L. *et al.* *Paleoceanography* **26**, PA4217 (2011).
9. Joughin, I., Smith, B. E. & Medley, B. *Science* <http://dx.doi.org/10.1126/science.1249055> (2014).
10. Rignot, E., Mouginot, J., Morlighem, M., Seroussi, H. & Scheuchl, B. *Geophys. Res. Lett.* <http://dx.doi.org/10.1002/2014GL060140> (2014).
11. Kopp, R. E., Simons, F. J., Mitrovica, J. X., Maloof, A. C. & Oppenheimer, M. *Nature* **462**, 863–867 (2009).
12. Dutton, A. & Lambeck, K. *Science* **337**, 216–219 (2012).

This article was published online on 28 May 2014.

## CELL BIOLOGY

# Short RNAs and shortness of breath

**The simultaneous deletion of six RNA molecules in mice has been found to cause respiratory and fertility defects, owing to improper assembly of structures called cilia on the cell surface. SEE ARTICLE P.115**

IRMA SÁNCHEZ & BRIAN D. DYNLACHT

**D**efects in the assembly of cilia — fine projections of the cell surface — are the cause of a plethora of human diseases<sup>1</sup>. One such disease is primary ciliary dyskinesia, a syndrome that severely compromises the function of the respiratory tract and reproductive system, resulting in airway infections and infertility. Identification of factors that control the assembly of cilia would constitute a major scientific advance. In this issue, Song *et al.*<sup>2</sup> (page 115) implicate clusters of non-protein-coding RNA molecules in the

regulation of cilium assembly, and show that deletion of these regulators in mice leads to a syndrome reminiscent of primary ciliary dyskinesia.

There are two types of cilium, non-motile and motile, both of which are essential for normal growth and development. Single, non-motile primary cilia are present on diverse cell types and function as sensors that provide the cell with information about the external environment. Motile cilia, meanwhile, function to coordinate fluid flow, and are found in large numbers on multiciliated cells (MCCs) in the respiratory tract and the Fallopian tubes

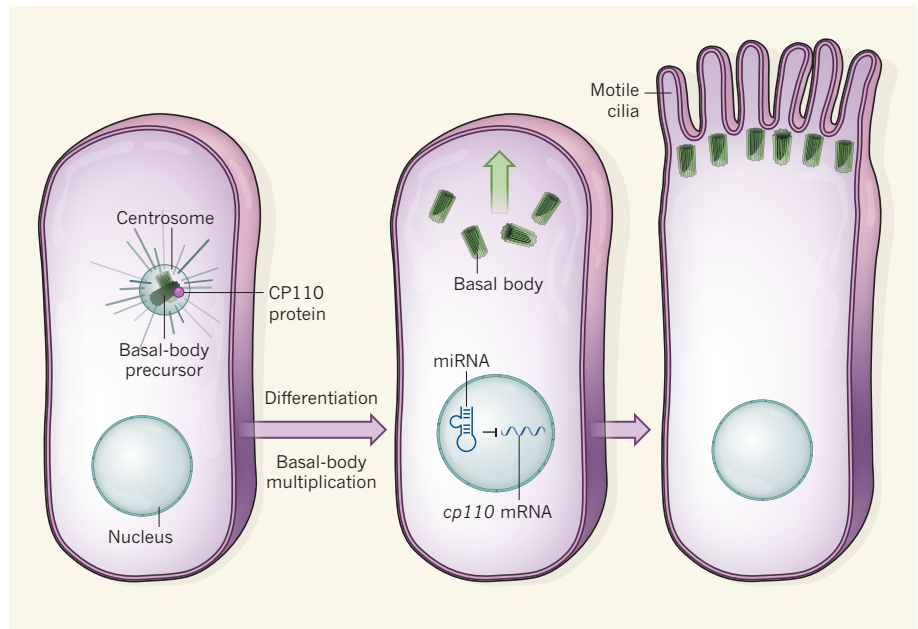
of the female reproductive tract. Assembly of both types (a process known as ciliogenesis) is initiated by a cellular organelle called the basal body. A basal-body precursor, located inside another organelle, the centrosome, matures and migrates to the plasma membrane, extending microtubular structures into the extracellular space to form the axoneme — the skeleton of the cilium (Fig. 1). But how this process is controlled is unclear.

A group of small, non-protein-coding RNAs called microRNAs (which regulate protein production by inhibiting the translation of messenger RNA) seem to be new-found regulators of motile ciliogenesis. The genetic region, or locus, known as miR-449 encodes three such miRNAs, and has been linked<sup>3,4</sup> to the development of a type of ciliated cell in the bronchi that is responsible for clearing mucus. A study<sup>5</sup> of the role of miR-449 in two types of MCC — the mucociliary epidermis cells of embryonic frogs and the cells that line the human bronchi — showed that expression from this locus increases as cellular precursors differentiate into MCCs. The miR-449 locus shares sequence similarity with two other miRNA loci, miR-34a and miR-34b/34c. Together, these three loci encode six miRNAs in vertebrates.

The miR-34/449 miRNAs are remarkably abundant in tissues enriched with motile cilia, comprising around 50% of the miRNAs expressed in the embryonic epidermis of frogs and 13% of the miRNAs in the cells that line the human respiratory tract<sup>5</sup>. To test whether the miR-34/449 cluster is essential for the initiation of ciliogenesis, Song *et al.* deleted all three loci in mice by mating existing strains harbouring single deletions. Simultaneous deletion was necessary because the six miRNA molecules have similar functions, and so can compensate for one another if only one locus is deleted. Triple-mutant mice failed to thrive — only 40% reached adulthood, and those that did were half the size of controls. The mice exhibited severe respiratory distress owing to a failure to clear mucus, and were sterile, mirroring defects observed in people with primary ciliary dyskinesia.

Song and colleagues obtained tracheal tissue from the mutant mice and examined it using high-speed imaging. They observed abnormal ciliation in mutant tracheal MCCs, which resulted in reduced fluid flow. This was not due to a developmental defect in the MCC precursors, but rather to a substantial reduction in the length and number of cilia per cell.

Multi-ciliation is achieved through an ordered process, beginning with the multiplication of basal bodies that subsequently dock at the cell surface. The authors found no differences in basal-body replication in mutant and control cells. Instead, high-resolution structural studies revealed that many basal bodies remained undocked in mutant cells, and so failed to assemble cilia. Those that did



**Figure 1 | Ciliogenesis.** The production of multiciliated cells from precursors occurs through an orderly process known as ciliogenesis, which begins with a single precursor of an organelle called the basal body, located inside another organelle, the centrosome. As the cell differentiates, the basal-body precursor multiplies and matures. Basal bodies then migrate and dock to the surface of the cell, extending microtubular structures into the extracellular environment to form cilia. Song *et al.*<sup>2</sup> report that, as precursors undergo differentiation, six miRNA molecules expressed by three related genetic regions regulate normal ciliogenesis in mice and frogs. The miRNAs inhibit translation of the messenger RNA that encodes the centrosomal protein CP110. The authors find that enhanced expression of CP110, brought about by simultaneous deletion of the six miRNAs, prevents normal docking of basal bodies to the cell surface (not shown).

dock generated significantly shorter, albeit structurally normal, axonemes.

Together, these data suggest that the miR-34/449 cluster might regulate basal-body docking. Song and co-workers screened a panel of genes for potential miR-34/449 targets, looking for genes that not only contained potential miR-34/449 binding sites but also showed downregulated expression during MCC differentiation (indicative of inhibition by highly expressed miRNAs). The authors identified more than 50 candidate genes from the screen, and chose the centrosomal protein CP110 as a focus for further investigation.

CP110 is an attractive target of miR-34/449 because previous work<sup>6</sup> has shown that overexpression of CP110 suppresses primary ciliary assembly, and that destruction of CP110 at basal bodies is required to trigger early stages of ciliogenesis. MicroRNAs are known to have multiple targets<sup>7</sup>, and the miR-34/449 cluster may regulate other transcripts that also influence ciliogenesis — much like miR-129-3p, which targets both *cp110* in ciliated cells and components of the cell's structural skeleton in other cell types<sup>8</sup>. However, when Song *et al.* inhibited *cp110* in frog embryos lacking miR-34/449 (which have increased levels of CP110), they found a substantial improvement in basal-body docking and fewer cilia defects, suggesting that excessive CP110 caused most of the observed defects. The current study, together with the previous report on

miR-129-3p (ref. 8), unifies a rapidly growing body of evidence establishing an evolutionarily conserved regulatory role for CP110 in cells with primary and motile cilia.

It is worth noting that proliferating and differentiating cells dedicate a remarkable level of attention to the control of CP110 abundance, which is regulated by transcriptional<sup>9</sup>, post-transcriptional<sup>2,3</sup> and post-translational mechanisms<sup>6</sup>. Indeed, no fewer than seven miRNAs (miR-129 and the miR-34/449 cluster)<sup>2,3,8</sup> and a pair of enzymes<sup>10,11</sup> have been identified as being involved in the robust regulation of *cp110* levels. It is likely that the findings of Song and colleagues represent the tip of the iceberg, and it will be exciting to explore the roles that other miRNAs have in modulating the levels of ciliary proteins that require similar fine tuning.

More generally, it will be useful to understand how subtle changes in the regulation of proteins such as CP110 could give rise to human disease. Intriguingly, links between aberrant CP110 levels and respiratory illness have already been uncovered in a previous study<sup>12</sup>. That study and the work of Song and colleagues indicate the importance of exploring the role of genes such as *cp110*, which regulate basal body and centrosomal components, in diseases associated with defects in ciliogenesis, including primary ciliary dyskinesia and human respiratory disease. ■



Irma Sánchez and Brian D. Dynlacht are at the New York University School of Medicine, New York, New York 10016, USA. e-mails: [irma.sanchez@med.nyu.edu](mailto:irma.sanchez@med.nyu.edu); [brian.dynlacht@med.nyu.edu](mailto:brian.dynlacht@med.nyu.edu)

1. Tsang, W. Y. & Dynlacht, B. D. *Cilia* **2**, 9 (2013).
2. Song, R. *et al. Nature* **510**, 115–120 (2014).

3. Lizé, M., Herr, C., Klimke, A., Bals, R. & Döbelstein, M. *Cell Cycle* **9**, 4579–4583 (2010).
4. Martínez-Anton, A. *et al. Am. J. Respir. Cell Mol. Biol.* **49**, 384–395 (2013).
5. Marcet, B. *et al. Nature Cell Biol.* **13**, 693–699 (2011).
6. Spektor, A., Tsang, W. Y., Khoo, D. & Dynlacht, B. D. *Cell* **130**, 678–690 (2007).
7. Flynt, A. S. & Lai, E. C. *Nature Rev. Genet.* **9**, 831–842 (2008).

8. Cao, J. *et al. Nature Cell Biol.* **14**, 697–706 (2012).
9. Chen, Z., Indjeian, V. B., McManus, M., Wang, L. & Dynlacht, B. D. *Dev. Cell* **3**, 339–350 (2002).
10. Li, J. *et al. Nature* **495**, 255–259 (2013).
11. D'Angiolella, V. *et al. Nature* **466**, 138–142 (2010).
12. Lai, Y. *et al. J. Allergy Clin. Immunol.* **128**, 1207–1215.e1 (2011).

## ASTROPHYSICS

# The MAD world of black holes

An analysis of optical and radio observations has revealed how powerful jets are launched from the centres of active galaxies, where supermassive black holes accrete matter through magnetically arrested disks, or MADs. [SEE LETTER P.126](#)

DENISE GABUZDA

Although the light given off by most galaxies is due to stars and glowing gas, some galaxies have extremely bright centres, or nuclei, with luminosities about 100,000 times greater than those of normal galaxies. In such active galactic nuclei, energy is liberated when matter spirals inwards and is captured by a supermassive black hole — billions of times more massive than the Sun — sitting at the galactic centre. In about 10% of these active nuclei, some of the in-spiralling matter is pushed into two jets of matter and radiation that shoot out in opposite directions at close to the speed of light (Fig. 1). However, the forces causing these outflows have remained unknown. In this issue, Zamaninasab *et al.*<sup>1</sup> (page 126) report direct evidence that the jets are launched on their

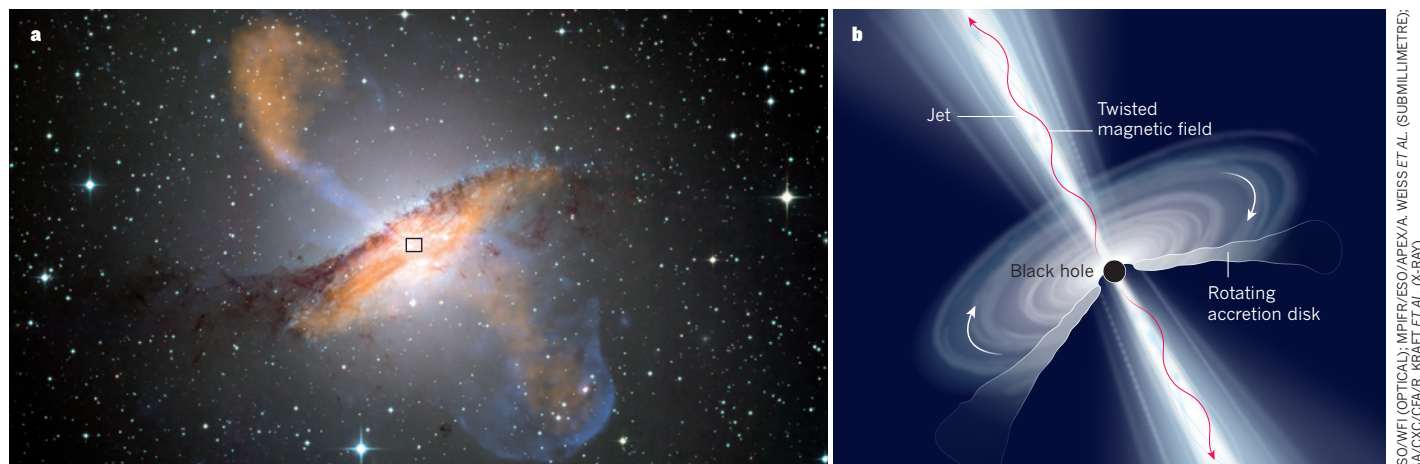
journey by a kind of gigantic electromagnetic generator, in which magnetic fields in the vicinity of the black hole are twisted by the black hole's spin, with the energy of this spin being transformed into the energy of the jets' outward motion<sup>2</sup>.

It has long been suspected that this mechanism might explain how jets in active galactic nuclei (AGN) are produced. However, direct observational evidence has been elusive, because the scales on which this mechanism should operate are tiny compared with the smallest scales that can be observed directly using the highest-resolution technique available, which are typically about 1 parsec ( $3 \times 10^{16}$  metres) from the central black hole. This scale might seem huge, but at the extremely large distances of AGN (billions of parsecs), it translates to a tiny angular distance as observed on the sky.

Images of such small scales can be obtained using very-long-baseline interferometry (VLBI), an elegant technique in which radio telescopes around the world observe in synchrony to imitate a single radio telescope with a diameter the size of Earth. The larger the telescope used, the finer the detail that can be seen; accordingly, VLBI yields radio images with phenomenally high resolution, equivalent to being able to peer across the Atlantic Ocean from the western edge of Europe and identify a coin held by someone standing on the eastern coast of the United States.

Unfortunately, however, the scales imaged with VLBI are still tens to thousands of times larger than those on which the powerful jets of AGN are launched. Zamaninasab *et al.* have found a clever way to bridge this gap, by considering the magnetic flux in the jets — essentially, the product of the magnetic field pointing along the jet and the jet's cross-section. The magnetic flux near the black hole cannot be measured directly, but should be proportional to the luminosity of the matter in the accretion disk surrounding the black hole. The disk's luminosity can be estimated from observations of optical lines in the spectrum of the AGN.

The authors have found a tight linear correlation between the estimated accretion-disk luminosities for 76 AGN and the magnetic fluxes in their jets measured with VLBI. This correlation demonstrates that the magnetic flux near the black hole is proportional to the



**Figure 1 | Jets emerging from the centre of an active galactic nucleus.** **a**, Composite image of the optical (true colour), X-ray (blue) and radio (orange) emission of the giant elliptical galaxy NGC 5128, showing its powerful radio jets extending far beyond the optical galaxy. **b**, Zamaninasab *et al.*<sup>1</sup> have found evidence that jets emerging from the centres of active galaxies are launched from deep in the galactic nucleus (square in **a**) by twisted magnetic fields<sup>2</sup>, and that these fields are dynamically important in the vicinity of the galaxy's central black hole and surrounding accretion disk.

magnetic flux far down the jets, as would be expected if the electromagnetic jet-launching mechanism referred to above is at work. The authors' results thus provide direct observational evidence that this is the case.

Theoretical simulations<sup>3,4</sup> of accretion disks have shown that, under certain conditions, the magnetic flux in the vicinity of the black hole naturally reaches a maximum equilibrium value. When this happens, forces exerted by the magnetic field dominate in the inner part of the disk. Disks for which this is true are called magnetically arrested disks, or MADs<sup>3,4</sup>. Zamaninasab *et al.* find that the derived slope of the linear correlation between the accretion-disk luminosity and the jet magnetic flux is precisely the value predicted for such disks, strongly suggesting that this MAD scenario is operating in the hearts of AGN.

These results indicate that the jets of AGN are launched electromagnetically by magnetic fields twisted by the black hole's spin, that these magnetic fields have a dominant role in determining the dynamics of the disk and jets in the vicinity of the central black hole, and that this may remain true at least out to VLBI scales, several parsecs from the black hole. It will therefore be important to consider the influence of the magnetic field, for example, when inferring the properties of the central black hole and accretion disk from high-resolution studies made with millimetre-wavelength, ground-based VLBI<sup>5–7</sup>, or with 'space VLBI', in which one or more antennas orbiting Earth are used with ground antennas<sup>8,9</sup>.

Zamaninasab and colleagues' findings also radically change the way astronomers view the jets emanating from the centres of AGN. These jets are not just outflows of matter carrying tremendous amounts of energy, but are also intrinsically magnetic structures. Many of their properties are probably determined by the magnetic fields embedded in them and travelling outwards with them. The twisting of the central magnetic fields that launches the jets should give rise to helical jet magnetic fields, which may be manifest in the jets' magnetic-field structure and morphology<sup>10,11</sup>. Because a fundamental relationship exists between magnetic fields and electrical currents, jet outflows should be regarded as systems of magnetic fields and currents. This is essential if we are to understand these enormous structures: how they propagate, why they remain so narrow as they traverse enormous distances, and how they interact with the material through which they are moving.

As the jets travel beyond their host galaxy and into intergalactic space, effects other than magnetic forces will probably also come into play, making the jets and the surrounding gas more turbulent and reducing the magnetic field's effects. Further detailed studies of the jets of AGN and their magnetic fields, from VLBI scales out to the ends of the jets many

thousands of parsecs from the central black hole, should help to determine whether such a transition occurs, and where. ■

**Denise Gabuzda** is in the Department of Physics, University College Cork, Cork, Ireland.  
e-mail: d.gabuzda@ucc.ie

1. Zamaninasab, M., Clausen-Brown, E., Savolainen, T. & Tchekhovskoy, A. *Nature* **510**, 126–128 (2014).
2. Blandford, R. D. & Znajek, R. L. *Mon. Not. R. Astron. Soc.* **179**, 433–456 (1977).
3. Tchekhovskoy, A., Narayan, R. & McKinney, J. C.

- Mon. Not. R. Astron. Soc.* **418**, L79–L83 (2011).
4. McKinney, J. C., Tchekhovskoy, A. & Blandford, R. D. *Mon. Not. R. Astron. Soc.* **423**, 3083–3117 (2012).
5. Dexter, J. & Fragile, P. C. *Mon. Not. R. Astron. Soc.* **432**, 2252–2272 (2013).
6. Doeleman, S. S. *et al. Science* **338**, 355–358 (2012).
7. Johannsen, T. *et al. Astrophys. J.* **758**, 30 (2012).
8. Kardashev, N. S. *et al. Astron. Rep.* **57**, 153–194 (2013).
9. Takahashi, R. & Mineshige, S. *Astrophys. J.* **729**, 86 (2011).
10. Molina, S. N. *et al. Astron. Astrophys.* (in the press); preprint at <http://arXiv.org/abs/1404.5961> (2014).
11. Gabuzda, D. C., Cantwell, T. M. & Cawthorne, T. V. *Mon. Not. R. Astron. Soc.* **438**, L1–L5 (2014).

## HEPATITIS C

# Treatment triumphs

**A stampede of recent clinical studies suggests that we are on the cusp of developing well-tolerated, orally delivered drugs that can effectively eradicate hepatitis C virus from most, if not all, infected individuals.**

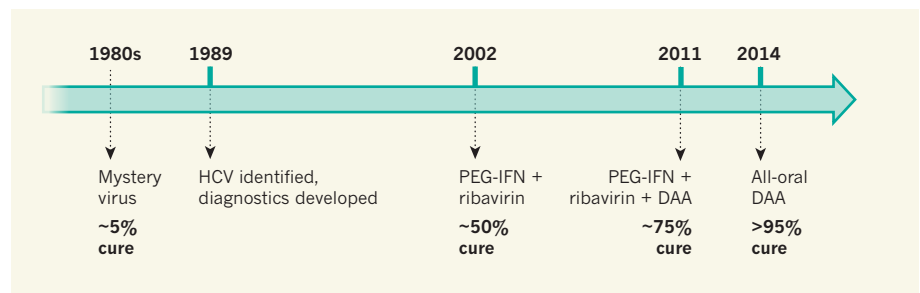
CHARLES M. RICE & MOHSAN SAEED

The story of hepatitis C began in the 1970s, when it was recognized that something other than hepatitis A or hepatitis B infections was causing liver inflammation following blood transfusions<sup>1,2</sup>. In 1989, the troublemaker was identified as a small RNA virus, named hepatitis C (HCV)<sup>3</sup>. Although there are now effective diagnostic procedures that allow a safe blood supply in most developed countries, intravenous drug abuse continues to lead to new infections. An estimated 185 million people are chronically infected with HCV and are at risk of developing life-threatening liver diseases, including cirrhosis and cancer<sup>4</sup>. But a recent series of clinical trials, reported in the *New England Journal of Medicine*<sup>5–11</sup>, demonstrate drastic increases in the effectiveness of anti-HCV drugs.

Historically, HCV-infected patients have

been treated by intravenous injections with type I interferons — secreted cellular proteins that elicit potent antiviral responses<sup>12</sup>. The success rates for interferon-based regimens improved from single digits in the 1970s to around 50% by 2002, accomplished by increasing dose, lengthening treatment, chemically stabilizing the interferon (by PEGylation) and adding ribavirin, an RNA-nucleoside analogue. Ribavirin has poor anti-HCV activity when used alone but significantly increased treatment success when combined with interferon (by mechanisms that are still unsettled). However, this treatment required a 24- or 48-week course and was plagued by awful side effects, including nausea, depression and anaemia. Hence, the goal remained to develop highly effective, orally administered and well-tolerated regimens that work for all patient groups.

Two enzymes encoded by HCV that are essential for viral replication — a serine



**Figure 1 | HCV trajectory.** In the 1980s, mysterious cases of liver inflammation following blood transfusions that were not explained by hepatitis A or hepatitis B viral infections were treated using type I interferon proteins, with a success rate of around 5%. The cause of these infections was identified in 1989 as RNA virus hepatitis C (HCV). The combination of PEGylated interferon (PEG-IFN) and ribavirin, approved in 2002, improved cure rates to around 50%. By 2011, drug cocktails containing HCV-specific direct-acting antivirals (DAAs) were being used to treat patients, with around 75% cure rates, and recent clinical trials<sup>5–11</sup> of all-oral, interferon-free, DAA-based regimens have increased treatment success rates to more than 95%.



protease (NS3-4A) and an RNA polymerase (NS5B) — are attractive drug targets. In the 2000s, inhibitors of these enzymes and of another non-enzymatic but essential HCV protein (NS5A), referred to as direct acting antivirals (DAAs), emerged as the lead targets for HCV drug development. In late 2011, two NS3-4A protease inhibitors were approved for human use in combination with PEGylated interferon and ribavirin, raising treatment success to more than 70% for patients with HCV genotype 1 (there are six highly divergent and variable genotypes of the virus).

However, euphoria over this advance was short-lived. Patients with advanced disease were treated but many others were not, owing to the additional, often severe, side effects of this drug combination and the emergence of viral resistance. In the meantime, and continuing into the present, dozens of new compounds were being tested in the clinic. In 2013, more-potent DAAs, in combination with PEGylated interferon and ribavirin, were approved, as was the first all-oral regimen, consisting of a NS5B-targeting DAA combined with ribavirin alone.

The recent clinical studies<sup>5–11</sup> present the next wave of interferon-free, all-oral, DAA-based regimens, which are likely to be approved in the near future for HCV treatment. Without delving into details and trade names, several key points about these trials emerge. First, they include multiple all-oral combinations that can achieve success rates of more than 95%. ‘Success’ for HCV treatment means no detectable virus 12 weeks after stopping treatment. Unlike drug treatments for hepatitis B and HIV, most HCV researchers believe that this endpoint represents a durable cure that lowers the risk of progressive liver disease. Second, these treatments are effective in patients who are in greatest need and are most difficult to treat — those with advanced fibrosis and cirrhosis, those who are co-infected with HIV, and even liver-transplant candidates and recipients. Also noteworthy is that the new drug combinations promise shorter treatment times (12 weeks and possibly even less) and minimal side effects; as a result, fewer people are expected to discontinue their treatment.

So from a mystery virus and a 5% treatment-success rate, we have come to an era of cure rates of more than 95% (Fig. 1). Game over, right? Not quite. What about viral resistance to the drugs? With nearly 200 million infected individuals, 6 diverse viral genotypes and around 1 trillion viral variants being generated per day per infected person, it is likely that HCV will have some tricks up its sleeve to develop resistance. However, some of the new DAAs, in particular sofosbuvir, which targets the active site of NS5B, have an extremely high barrier to resistance, and there have been only rare glimpses of resistant variants in clinical observations with multiple viral genotypes<sup>13</sup>. Combining potent DAAs, each with lower resistance barriers, can still be highly

effective at avoiding the build-up of resistance. Nonetheless, resistance will undoubtedly occur and should be taken into account to guide treatment decisions. The current drugs are also less effective against genotype 3 HCV, which is common in South Asia, although pan-genotype drugs are in development.

Another barrier is identifying those infected. Most people are unaware of their HCV infection<sup>14</sup>, and only a small minority has been treated<sup>15</sup>. Although some health agencies have recommended universal screening of high-risk groups, implementing such policies is challenging and time-consuming. And once infected individuals are identified, how will society pay for their treatment? The current price tag for cutting-edge HCV treatment in the United States is more than US\$80,000 for a 12-week course. Competition among pharmaceutical companies may lower this price, but most people infected with HCV live in countries that cannot afford the new treatments. Fortunately, there is movement in the pharmaceutical industry to provide for low-cost drug production in certain countries, such as Egypt, where an estimated 10% of the population is infected. Finally, getting rid of the virus does not always erase the risk of future liver-related problems — patients still need to be monitored routinely for liver function and cancer, particularly those whose infection had led to cirrhosis.

With the new drugs that are in hand or on the horizon, we have the means to eradicate this virus, possibly without needing a vaccine. However, the challenge now is to extend these

great medical advances on a national and global scale to those in need — something that has not been terribly effective in the past. We can hope that implementing these transformative HCV advances will help to create a model for success, for this and other widespread human diseases. ■

**Charles M. Rice and Mohsan Saeed** are in the Center for the Study of Hepatitis C, Laboratory of Virology and Infectious Disease, The Rockefeller University, New York, New York 10065, USA.  
e-mail: ricec@rockefeller.edu

1. Prince, A. M. et al. *Lancet* **2**, 241–246 (1974).
2. Alter, H. J. et al. *Lancet* **2**, 838–841 (1975).
3. Choo, Q. L. et al. *Science* **244**, 359–362 (1989).
4. Mohd Hanafiah, K., Groeger, J., Flaxman, A. D. & Wiersma, S. T. *Hepatology* **57**, 1333–1342 (2013).
5. Feld, J. J. et al. *N. Engl. J. Med.* **370**, 1594–1603 (2014).
6. Afdhal, N. et al. *N. Engl. J. Med.* **370**, 1889–1898 (2014).
7. Afdhal, N. et al. *N. Engl. J. Med.* **370**, 1483–1493 (2014).
8. Kowdley, K. V. et al. *N. Engl. J. Med.* **370**, 222–232 (2014).
9. Kowdley, K. V. et al. *N. Engl. J. Med.* **370**, 1879–1888 (2014).
10. Zeuzem, S. et al. *N. Engl. J. Med.* **370**, 1604–1614 (2014).
11. Sulkowski, M. S., Jacobson, I. M. & Nelson, D. R. *N. Engl. J. Med.* **370**, 1560–1561 (2014).
12. Heim, M. H. *Nature Rev. Immunol.* **13**, 535–542 (2013).
13. Lawitz, E. et al. *N. Engl. J. Med.* **368**, 1878–1887 (2013).
14. Denniston, M. M., Kleven, R. M., McQuillan, G. M. & Jiles, R. B. *Hepatology* **55**, 1652–1661 (2012).
15. Dore, G. J., Ward, J. & Thursz, M. J. *Viral Hepat.* **21** (suppl. 1) 1–4 (2014).

#### NEUROLOGICAL DISORDERS

## Quality-control pathway unlocked

**A modified ubiquitin protein has been identified by three independent studies as the missing link in a cellular quality-control pathway that is implicated in Parkinson's disease. SEE LETTER P.162**

ASA ABELIOVICH

**P**arkinson's disease, a progressive neurodegenerative disorder, has long been hypothesized to be caused by defects in organelles called mitochondria, which power mammalian cells through the production of ATP molecules. An accumulation of dysfunctional mitochondria may lead not only to a cellular energy crisis, but also to excessive production of toxic by-products. Two enzymes implicated in Parkinson's disease, PINK1 and parkin<sup>1,2</sup>, are thought to be involved in the disposal of defective mitochondria, but how

the two proteins interact has been unclear. A trio of studies (by Kane *et al.*<sup>3</sup>, writing in the *Journal of Cell Biology*; by Kazlauskaitė *et al.*<sup>4</sup>, in the *Biochemical Journal*; and by Koyano *et al.*<sup>5</sup>, on page 162 of this issue) now report that phosphorylated ubiquitin protein is the link between PINK1 and parkin, providing insights into a complex system of parkin regulation.

Kinase enzymes such as PINK1 alter the behaviour of target proteins through the addition of phosphate groups, a process called phosphorylation. PINK1 is imported to mitochondria and, in healthy cells, undergoes

rapid degradation<sup>6</sup>. However, if mitochondria are defective or damaged (for example by exposure to CCCP, a poison that blocks ATP production), PINK1 accumulates, becoming anchored to the outer mitochondrial membrane with its kinase domain exposed to the cytoplasm.

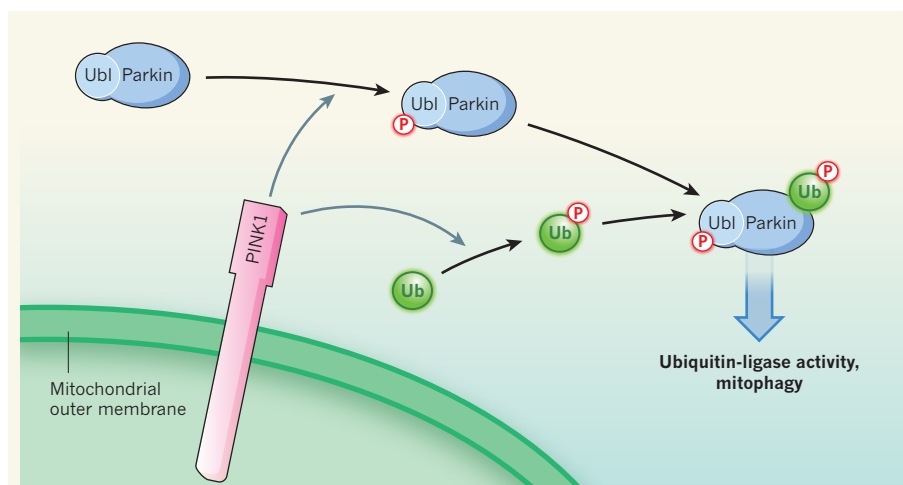
Damaged mitochondria also attract parkin, which is otherwise dispersed throughout the cytoplasm in healthy cells<sup>7</sup>. Parkin is a ubiquitin ligase, which adds ubiquitin proteins (either singly or in polyubiquitin chains) both to itself through autoubiquitination and to nearby target proteins. Ubiquitinated proteins can serve as a signal to the cell that a cellular compartment should be degraded, which in damaged mitochondria leads to their timely disposal<sup>7</sup>, a process known as mitophagy.

Mutations in either *PINK1* or *PARKIN* that underlie rare familial forms of Parkinson's disease disrupt mitophagy, implicating this cellular pathway in Parkinson's disease<sup>7</sup>. Furthermore, *PINK1* mutations impede the recruitment of parkin to damaged mitochondria, suggesting that the proteins act in a linear pathway. Consistent with a PINK1–parkin quality-control pathway, mutations in *pink1* or *parkin* in fruit flies cause accumulation of defective mitochondria and cellular degeneration<sup>8,9</sup>.

Initial models proposed that PINK1 phosphorylates and so activates parkin in damaged mitochondria. Although direct phosphorylation of parkin by PINK1 has been documented<sup>10</sup>, this modification does not seem to be sufficient for full activation of parkin's ubiquitin-ligase activity<sup>3–5,10</sup>. In search of a functional connection between PINK1 and parkin, three groups undertook cell-wide protein analyses and biochemical studies, and found the missing link between the two — phosphorylated ubiquitin (phospho-ubiquitin).

Each study showed that, in cells in which PINK1 was activated by CCCP treatment, PINK1 phosphorylates ubiquitin at a serine amino-acid residue (serine 65). Strikingly, a corresponding serine-65 residue in a ubiquitin-like domain is the aforementioned target of PINK1 phosphorylation on parkin<sup>10</sup>. Subsequent analyses by all three groups demonstrated that modified ubiquitin, in turn, induces parkin activity (Fig. 1).

Koyano and co-workers found that modified ubiquitin alone could not fully activate parkin — complete activation required coincident modification of parkin's ubiquitin-like domain as well as of ubiquitin, each at their respective serine-65 residues. A unique aspect of this group's work is their use of a strain of yeast that harbours a mutant form of ubiquitin lacking the serine-65 residue, which cannot be phosphorylated by PINK1. When the authors added human PINK1 and parkin to these cells, they found that parkin was not activated, underscoring the idea of an ordered pathway for mitophagy.



**Figure 1 | PINK1 and parkin in mitochondrial quality control.** Mitochondrial damage leads to anchoring of the PINK1 enzyme to the outer mitochondrial membrane, with its kinase domain facing the cytoplasm. PINK1 adds a phosphate group (P) to the ubiquitin-like domain (Ubl) of the ubiquitin-ligase enzyme parkin. Three studies<sup>3–5</sup> find that PINK1 also phosphorylates the ubiquitin (Ub) protein itself. Phosphorylated ubiquitin directly binds to and activates parkin. Activated parkin ligates ubiquitin and phospho-ubiquitin molecules to nearby target proteins, leading to disposal of the damaged mitochondria through mitophagy.

Whereas all three studies implicate phosphorylated ubiquitin as an intermediary in the PINK1–parkin pathway, the role of direct phosphorylation of parkin by PINK1 seems more complex. Koyano and colleagues report that modification of both ubiquitin and parkin at serine-65 is necessary for full activation of parkin in cells. But Kane and colleagues found evidence that modification of ubiquitin alone can activate parkin. This discrepancy is likely to relate to the distinct assays used in the studies, rather than to a biological difference.

Consistent with phospho-ubiquitin's activating role, Kane *et al.* and Koyano *et al.* found that it binds directly to parkin. Koyano and colleagues took the studies a step further, demonstrating that phospho-ubiquitin can still be used by parkin as a substrate for ubiquitination and autoubiquitination. But, surprisingly, the group found that parkin could be activated by phospho-ubiquitin that was mutated or modified such that it could not act directly as a substrate in ubiquitination. This implies that phospho-ubiquitin binds to and activates parkin separately from its role as a substrate.

Clues as to how this could be achieved might be gleaned from recent crystallographic analyses of parkin<sup>11,12</sup>. A phospho-peptide binding pocket has been proposed<sup>11</sup> to lie within an inhibitory domain in parkin that, when the protein is inactive, occludes access to its catalytic active site. Kazlauskaitė *et al.* speculate that the active site of parkin could be exposed by conformational changes brought about by the binding of phospho-ubiquitin's phosphate group to this inhibitory domain.

Kane and co-workers' data point to another role for phospho-ubiquitin — recruiting parkin to the outer membrane of damaged mitochondria. A particularly interesting idea is

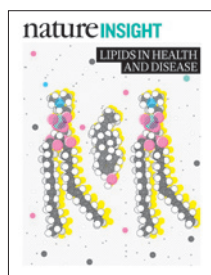
that such recruitment may generate a positive feedback loop, in which recruited parkin would be predicted to ligate additional phospho-ubiquitin to nearby proteins, attracting yet more parkin.

A subset of known parkin substrates, including the proteins mitofusin 2 and Miro, regulate mitochondria<sup>13,14</sup>, and their ubiquitination by parkin may be required for normal mitophagy. It will be important to determine whether activation by phospho-ubiquitin affects parkin's target selection, the fate of ubiquitinated target proteins, or the structure of polyubiquitin chains formed on targets. Finally, drugs that mimic the effects of phospho-ubiquitin may be candidate therapeutics for inherited and sporadic forms of Parkinson's disease. ■

**Asa Abeliovich** is in the Departments of Pathology, Cell Biology and Neurology, and at the Taub Institute, Columbia University, New York, New York 10032, USA. e-mail: aa900@columbia.edu

- Valente, E. M. *et al.* *Science* **304**, 1158–1160 (2004).
- Kitada, T. *et al.* *Nature* **392**, 605–608 (1998).
- Kane, L. A. *et al.* *J. Cell Biol.* **205**, 143–153 (2014).
- Kazlauskaitė, A. *et al.* *Biochem. J.* **460**, 127–139 (2014).
- Koyano, F. *et al.* *Nature* **510**, 162–166 (2014).
- Jin, S. M. *et al.* *J. Cell Biol.* **191**, 933–942 (2010).
- Narendra, D. P. *et al.* *PLoS Biol.* **8**, e1000298 (2010).
- Clark, I. E. *et al.* *Nature* **441**, 1162–1166 (2006).
- Park, J. *et al.* *Nature* **441**, 1157–1161 (2006).
- Shiba-Fukushima, K. *et al.* *Sci. Rep.* **2**, 1002 (2012).
- Wauer, T. & Komander, D. *EMBO J.* **32**, 2099–2112 (2013).
- Trempe, J. F. *et al.* *Science* **340**, 1451–1455 (2013).
- Poole, A. C. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 1638–1643 (2008).
- Ziviani, E., Tao, R. N. & Whitworth, A. J. *Proc. Natl Acad. Sci. USA* **107**, 5018–5023 (2010).





**Cover illustration**  
Jasiek Krzysztofciak

**Editor, *Nature***  
Philip Campbell

**Publishing**  
Richard Hughes

**Production Editor**  
Jenny Rooke

**Art Editor**  
Nik Spencer

**Sponsorship**  
Reya Silao

**Production**  
Ian Pope

**Marketing**  
Elena Woodstock  
Steven Hurst

**Editorial Assistant**  
Abbie Williams

The Macmillan Building  
4 Crinan Street  
London N1 9XW, UK  
Tel: +44 (0) 20 7833 4000  
e: nature@nature.com



nature publishing group

**L**ipids are small hydrophobic molecules that carry out a multitude of crucial roles. For example, they act as structural elements in biological membranes, they store energy and they function as signalling molecules in cellular response pathways. Disruption to their levels of expression — as occurs in obesity, diabetes, autoimmunity or inborn errors of lipid metabolism — leads to dysfunction and disease in many organs.

In this Insight, Joost Holthuis and Anant Menon begin by discussing the cellular pathways by which eukaryotic cells maintain a variety of lipid compositions in different organelles so that they can carry out their specialized tasks.

Next, Michael Maceyka and Sarah Spiegel discuss the molecular mechanisms by which the metabolites of sphingolipids exert their cell signalling and pathological functions.

Sphingolipid lysosomal storage disorders are characterized by the accumulation of glycosphingolipids in the late endocytic system. Frances Platt discusses insights into the cell biology of glycosphingolipid storage disorders and the development of treatments.

Vivian Peirce and colleagues then go on to highlight the biology of various types of adipose tissues, and their therapeutic potential in metabolic disease.

The Perspective by Gerald Shulman and colleagues discusses the roles of hepatic lipids in non-alcoholic fatty liver disease, which is now the most common liver disorder in the Western world.

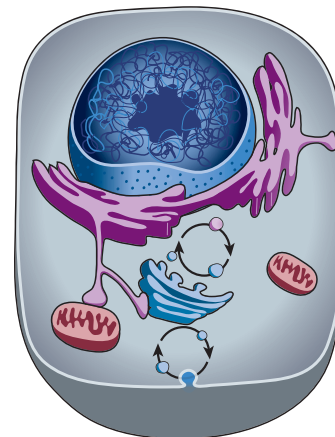
Finally, Charles Serhan reviews the anti-inflammatory and immunoregulatory actions of omega-3-derived lipid resolving mediators and their therapeutic potential.

**Joshua Finkelstein, Marie-Thérèse Heemels,  
Sadaf Shadan & Ursula Weiss**  
*Senior Editors*

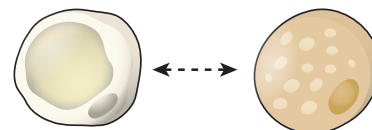
### CONTENTS

#### REVIEWS

- 48 Lipid landscapes and pipelines in membrane homeostasis**  
*Joost C. M. Holthuis & Anant K. Menon*



- 58 Sphingolipid metabolites in inflammatory disease**  
*Michael Maceyka & Sarah Spiegel*
- 68 Sphingolipid lysosomal storage disorders**  
*Frances M. Platt*
- 76 The different shades of fat**  
*Vivian Peirce, Stefania Carobbio & Antonio Vidal-Puig*



#### PERSPECTIVE

- 84 The role of hepatic lipids in hepatic insulin resistance and type 2 diabetes**  
*Rachel J. Perry, Varman T. Samuel, Kitt F. Petersen & Gerald I. Shulman*

#### REVIEW

- 92 Pro-resolving lipid mediators are leads for resolution physiology**  
*Charles N. Serhan*

# Lipid landscapes and pipelines in membrane homeostasis

Joost C.M. Holthuis<sup>1,2</sup> & Anant K. Menon<sup>3</sup>

**The lipid composition of cellular organelles is tailored to suit their specialized tasks. A fundamental transition in the lipid landscape divides the secretory pathway in early and late membrane territories, allowing an adaptation from biogenic to barrier functions. Defending the contrasting features of these territories against erosion by vesicular traffic poses a major logistical problem. To this end, cells evolved a network of lipid composition sensors and pipelines along which lipids are moved by non-vesicular mechanisms. We review recent insights into the molecular basis of this regulatory network and consider examples in which malfunction of its components leads to system failure and disease.**

The propensity of the non-polar acyl chains of lipids to self-associate in water is the physical basis for spontaneous membrane formation, and renders membranes virtually impermeable to polar solutes. This fundamental principle enabled the first cells to segregate their internal constituents from the external environment and is repeated within nucleated cells to form discrete organelles. Whereas a single lipid species is sufficient to generate a barrier to unregulated solute movement, membrane lipids in eukaryotic cells are characterized by a remarkable structural diversity<sup>1,2</sup>. In addition, the lipid composition and transbilayer arrangement among organelles show striking variations, and there is compelling evidence that the collective properties of the bulk lipids play a profound part in defining organelle identity and function.

Of particular interest are the lipid-induced changes in physical membrane properties such as bilayer thickness, lipid packing density and surface charge that mark the transition from early to late organelles in the secretory pathway<sup>3,4</sup>. These alterations are conserved from yeast to human and provide specific cues for membrane proteins to allow temporal and spatial control over the molecular events that govern secretion. This requires not only elaborate metabolic machinery for creating a diverse repertoire of membrane lipids, but also ways to preserve the unique lipid mixtures of early and late secretory organelles. Without mechanisms imposing selectivity and directionality on lipid movement, organelle identity and function would vanish in the face of extensive vesicular trafficking.

In this Review, we discuss how cells coordinate lipid metabolism and transport to achieve membrane lipid homeostasis in organelles connected by vesicular trafficking. As the underlying molecular principles involve collective phenomena beyond specific interactions among proteins and lipids, we first describe how lipid composition influences the generic physical properties of the membrane and how these properties change from early to late secretory organelles. We discuss how these changes are monitored by membrane sensors that relay signals to the lipid metabolic machinery when organelle lipid homeostasis is challenged. We then focus on the long-standing enigma that lipid traffic in cells is largely independent of membrane traffic. Current models suggest that lipid traffic is mediated by cytoplasmic transfer proteins that operate at contact sites between early and late secretory organelles. An appealing concept is that these so-called lipid pipelines<sup>5</sup> are exploited to bypass vesicular connections for optimal

control over lipid exchange between secretory organelles. Finally, we discuss mounting evidence of extensive co-regulation among individual lipid classes and how perturbation of the underlying homeostatic machinery can culminate in system failure and disease.

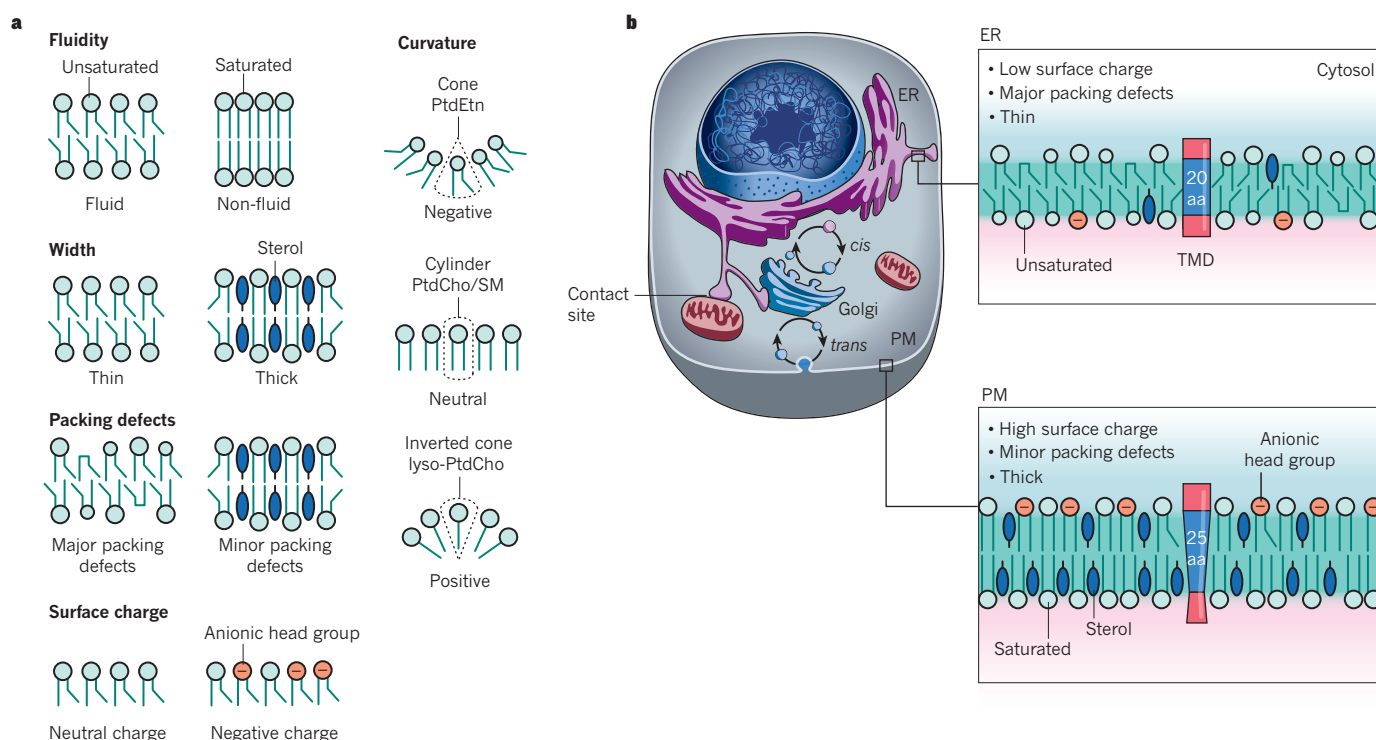
## Lipid repertoire of the secretory pathway

The organelles along the secretory pathway have major differences in lipid composition that help to shape their specialized tasks. Whereas some rare lipids (for example, phosphoinositides) contribute to organelle identity by mediating specific molecular recognition events, we focus on how the collective action of the bulk lipids specify organelle function by influencing generic physical membrane parameters such as fluidity, thickness, lipid packing density and surface charge (Fig. 1a). For proper organelle function, these parameters must be kept within an appropriate range. To better understand the underlying principles, we first discuss some physico-chemical properties of the main categories of membrane lipids found along the secretory pathway — phospholipids, sphingolipids and sterols.

Phospholipids constitute the bulk of the membrane's lipid matrix. They contain two hydrophobic acyl chains and a phosphate head group ester linked to a glycerol backbone. Further classification is based on the structure of the phospholipid head group, which is zwitterionic in phosphatidylcholine (PtdCho) and phosphatidylethanolamine (PtdEtn), or anionic in phosphatidylserine (PtdSer) and phosphatidylinositol (PtdIns). Each phospholipid class comprises a multitude of molecular species defined by the length and degree of saturation in their acyl chains. Both head group and acyl chain composition influence the physical properties of the membrane<sup>6</sup>. PtdCho accounts for more than 50% of all phospholipids in most eukaryotes and usually contains one *cis*-unsaturated acyl chain, such as oleic acid (C18:1). The rigid kink of the *cis*-double bond lowers the packing density of the acyl chains, which increases membrane fluidity<sup>7</sup>. Facilitated by its cylindrical shape, PtdCho self-assembles spontaneously into closed bilayers that adopt a liquid crystalline state at physiological temperature. Thus, PtdCho is ideally suited to providing a stable and fluid matrix for cellular membranes, a pre-condition for the living state. By contrast, PtdEtn is a conically shaped phospholipid that imposes negative curvature stress on the membrane owing to its relatively small polar head group. The non-bilayer propensity of PtdEtn increases with acyl chain unsaturation, creating lipid-packing

<sup>1</sup>Fachbereich Biologie/Chemie, University of Osnabrück, Barbarastrasse 13, 49076 Osnabrück, Germany. <sup>2</sup>Membrane Biochemistry & Biophysics, Bijvoet Center for Biomolecular Research and Institute of Biomembranes, Utrecht University, the Netherlands. <sup>3</sup>Department of Biochemistry, Weill Cornell Medical College, 1300 York Avenue, New York, NY 10065, USA.





**Figure 1 | Membranes of early and late secretory organelles have contrasting lipid compositions and physical properties.** **a**, Physical membrane properties are influenced by lipid composition. Fluidity is promoted by lipids with short, unsaturated fatty acids. The double bonds introduce kinks that lower the packing density of the acyl chains and inhibit transition of the membrane from a fluid to a solid gel phase. Phospholipids typically pair a saturated with an unsaturated fatty acid, which is an elegant means to prevent the two types of acyl chains from phase separating when the temperature drops. Thickness is promoted by acyl-chain length and sterols, which order and stretch the acyl chains. High levels of packing defects are found in lipids with unsaturated acyl chains and small head groups. Surface charge is determined by the presence of anionic lipids such as phosphatidylserine and phosphoinositides. Curvature is determined by lipid shape. Lipids with a small area ratio of polar head to acyl chain (creating a cone shape) induce negative curvature, lipids with an equal head to chain ratio (creating a cylinder shape)

defects that facilitate membrane fusion and influence the binding and activity of peripheral membrane proteins<sup>8,9</sup>. Despite their relatively low abundance, PtdSer and PtdIns are key determinants of membrane surface charge and mediate functional interactions with positively charged regions of peripheral and integral membrane proteins<sup>10,11</sup>.

Distinct from phospholipids, sphingolipids primarily contain saturated or *trans*-unsaturated acyl chains linked to a serine backbone. This hydrophobic structure, termed ceramide, is decorated with a phosphocholine head group in the case of sphingomyelin or with saccharides in the case of glycosphingolipids. Owing to its acyl-chain composition, sphingomyelin forms a taller, narrower cylinder than PtdCho, increasing its packing density in the membrane. Consequently, at physiological temperatures a sphingomyelin bilayer exists in a solid gel phase with tightly packed, immobile acyl chains<sup>12–14</sup>. However, the addition of sterols renders these membranes fluid<sup>15</sup>. Sterols are abundant apolar membrane lipids with an inflexible core formed by four fused rings, with cholesterol predominating in mammals. By interfering with acyl-chain packing, sterols inhibit the transition of the membrane to the solid gel state. At the same time, sterols rigidify fluid membranes by reducing the flexibility of neighbouring unsaturated acyl chains, thereby increasing membrane thickness and impermeability to solutes (the so-called condensing effect of sterols)<sup>15,16</sup>. Thus, sterols allow cells to minimize unregulated solute movement across their membranes while keeping them fluid over a

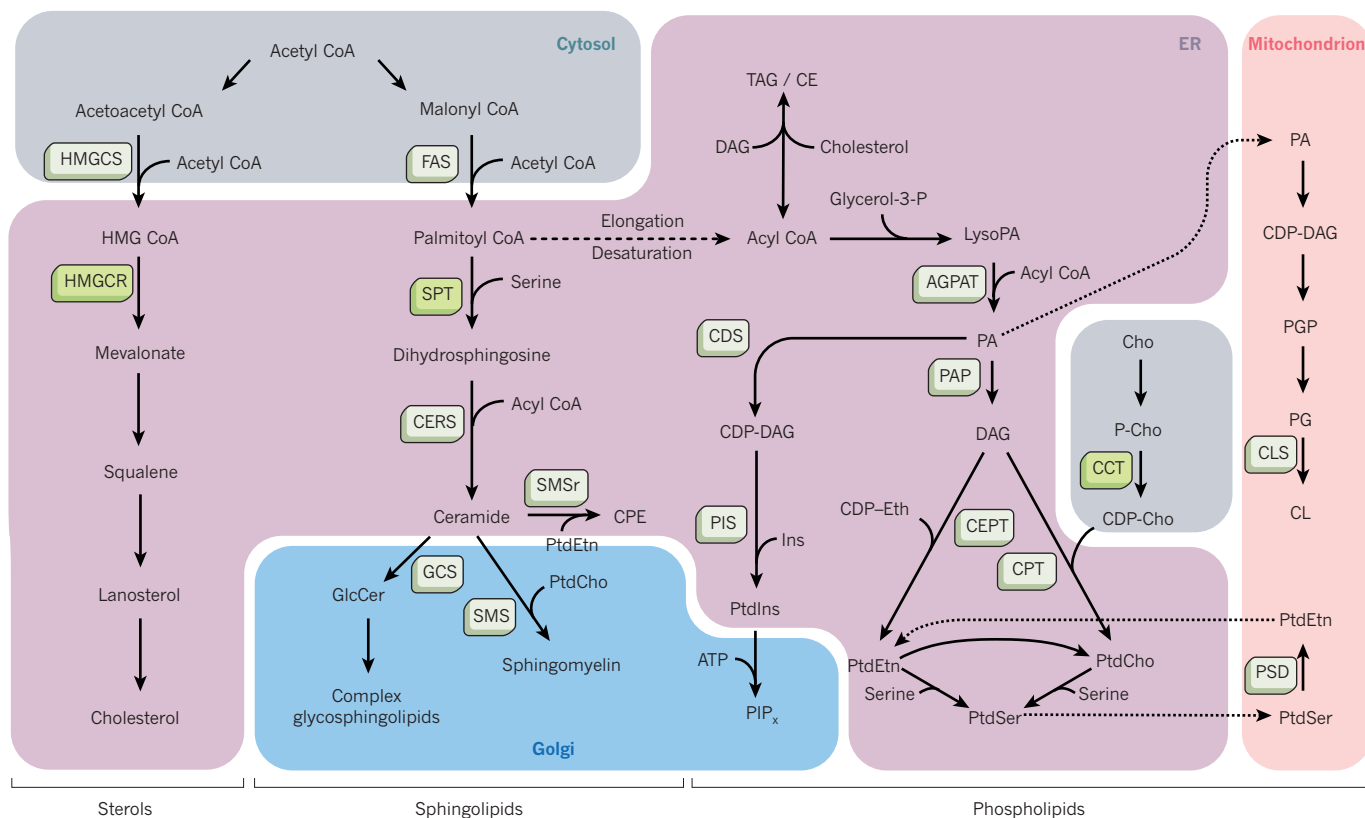
are neutral, and those with a much larger head compared with the acyl chain area (creating an inverted cone shape) induce a positive curvature. PtdCho, phosphatidylcholine; PtdEtn, phosphatidylethanolamine; SM, sphingomyelin. **b**, The endoplasmic reticulum (ER) has a thin bilayer, loose lipid packing and neutral cytoplasmic surface charge adapted for its biogenic function. The plasma membrane (PM) has a thick bilayer, tight lipid packing and negative cytoplasmic surface charge adapted for its barrier function. These contrasting physical properties are reflected in the length (20 aa and 25 aa, respectively) and geometry of transmembrane domains (TMDs) of ER and plasma-membrane-resident proteins. Membrane traffic between the ER and plasma membrane passes through the Golgi, a polarized and multi-cisternal organelle in which lipid sorting must occur to preserve the unique lipid compositions and, hence, specialized functions of the ER and plasma membrane. Note that lipid traffic can bypass membrane traffic at sites at which the ER contacts the *trans*-Golgi, plasma membrane or mitochondria.

broad range of acyl-chain compositions and temperatures.

Restriction of the membrane lipid repertoire of eukaryotic cells is not well tolerated. Attempts to manipulate the ratio between cylindrical and conical lipids indicate that this ratio is tightly controlled through compensatory mechanisms<sup>17</sup>. Cells defective in PtdSer biosynthesis preserve membrane electrostatics by upregulating the PtdIns content of their membranes<sup>18</sup>. The requirement of sphingolipids for cell growth in yeast can be bypassed by a suppressor mutation that enables the production of a novel set of phospholipids, the structural and physico-chemical properties of which mimic those of the sphingolipids<sup>19</sup>. We discuss how eukaryotic cells exploit a structurally diverse set of membrane lipids to sustain the advanced specialization of their secretory apparatus.

### Lipid composition subserves secretory organelle function

The endoplasmic reticulum (ER) is the gateway of the secretory pathway and the principal site for membrane protein insertion and folding<sup>20</sup>. It is also the cell's main lipid factory, in which the bulk of phospholipids and sterols, as well as substantial amounts of storage lipids such as triacylglycerol and steryl esters, are produced (Fig. 2). In addition, the ER synthesizes ceramide, the precursor of all sphingolipids. Besides exporting ceramide, the ER supplies a large portion of membrane lipids to the Golgi and plasma membrane because these distal secretory organelles have little or no capacity to produce their



**Figure 2 | Membrane lipid biosynthesis.** Principal pathways for the production of sterols, sphingolipids and phospholipids in mammals, and the key biosynthetic enzymes (pale green) and rate-limiting enzymes (green) involved. AGPAT, 1-acylglycerol-3-phosphate-O-acyltransferase; CCT, CTP:phosphocholine cytidyltransferase; CDP-Cho, cytidinediphosphate choline; CDP-DAG, cytidine diphosphate-diacylglycerol; CDP-Eth, cytidine diphosphate ethanolamine; CDS, cytidine diphosphate diacylglycerol synthase; CEPT, choline/ethanolaminephosphotransferase; CERS, ceramide synthase; Cho, choline; CL, cardiolipin; CLS, cardiolipin synthase; CPE, ceramide phosphoethanolamine; CPT, cholinephosphotransferase;

FAS, fatty acid synthase; GCS, glucosylceramide synthase; HMGCS, hydroxymethylglutaryl coenzyme A (CoA) synthase; HMGCR, 3-hydroxy-3-methylglutaryl-CoA reductase; Ins, inositol; PA, phosphatidic acid; PAP (also known as LIPIN), phosphatidic acid phosphatase; P-Cho, phosphocholine; PG, phosphatidylglycerol; PGP, phosphatidylglycerolphosphate; PIP, phosphoinositide; PtdCho, phosphatidylcholine; PtdEtn, phosphatidylethanolamine; PtdInsS, phosphatidylinositol synthase; PtdSer, phosphatidylserine; SMS, sphingomyelin synthase; SMSr, sphingomyelin synthase-related enzyme; SPT, serine palmitoyltransferase; TAG/CE, triacylglycerol/cholesterol ester.

own. Despite an extensive exchange of material by membrane trafficking, the ER and plasma membrane show remarkable differences in their lipid composition (Fig. 1b). For instance, sterols are rare in the ER (5 mol% of lipids) but abundant in the *trans*-Golgi and plasma membrane (30–40 mol%)<sup>12,21</sup>. Sterols preferentially interact with lipids bearing saturated fatty acyl chains and bulky head groups, such as sphingolipids<sup>13</sup>. The bulk of sphingolipids are synthesized in the lumen of the *trans*-Golgi from ceramide supplied by the ER<sup>22</sup>. Analogous to sterols, sphingolipids accumulate in the plasma membrane while their levels are low in the ER. Moreover, phospholipids are more saturated at the plasma membrane than at the ER<sup>23</sup> owing to a substantial remodelling of their acyl chains<sup>6</sup>. Thus, the plasma membrane is characterized by tight lipid packing due to a high concentration of saturated lipid species and sterols, making it a rigid and thick barrier between the intracellular and external milieu.

The transition from a thin and loosely packed membrane into a thick and rigid one occurs in the *trans*-Golgi and is driven by sphingolipid production and sterol supply<sup>22</sup> (discussed later). The mechanism by which sphingolipids and sterols are moved up a concentration gradient into the plasma membrane is incompletely understood, but their preferential interactions and co-depletion from the highly curved coat protein complex 1 (COPI) vesicles that mediate retrograde membrane trafficking might have a role<sup>22,24,25</sup>. In any case, the step-change in membrane thickness caused by the abrupt increase in the concentration of sterols and sphingolipids is reflected in the length and composition of transmembrane domains (TMDs)

of single-spanning membrane proteins that populate the secretory pathway. TMDs of ER and Golgi-resident proteins are significantly shorter than those of plasma-membrane proteins and have more bulky residues in the exoplasmic hemi-TMD<sup>3</sup> (Fig. 1b). Both TMD length and volume are crucial determinants of Golgi retention or exit to the plasma membrane<sup>26,27</sup>. Selective associations between matching lipids and TMDs can drive their co-segregation in model membranes<sup>28</sup>. This implies that the Golgi exploits physical principles such as hydrophobic mismatch and packing density to solve a complex logistical problem, namely to preserve the unique mixtures of membrane proteins and lipids that allow the ER and plasma membrane to execute their fundamentally distinct tasks.

Recent work indicates that the Golgi also marks a transition point in membrane electrostatics. Membrane surface charge largely depends on the content of anionic phospholipids such as PtdSer, the levels of which range from a few per cent in the ER to more than 10% at the plasma membrane<sup>12</sup>. Studies with genetically encoded PtdSer probes indicate that the bulk of PtdSer resides in the luminal leaflet of the ER, an asymmetric distribution opposite to that of the plasma membrane in which PtdSer faces the cytoplasm<sup>29</sup>. Bulk phospholipids, including PtdSer, are synthesized in the cytoplasmic leaflet of the ER and then randomly 'scrambled' by a bidirectional flippase to allow membrane expansion<sup>30</sup>. This suggests that PtdSer is trapped in the ER lumen, possibly through calcium-mediated interactions with luminal proteins. PtdSer asymmetry in the ER is consistent with the finding that the phospholipid monolayer around lipid droplets,



which is thought to arise from the cytoplasmic leaflet of the ER, contains hardly any PtdSer<sup>31</sup>. Most of the Golgi does not react with the cytoplasmic PtdSer probe, except for the *trans*-Golgi, which contains type 4 P-type ATPases (P<sub>4</sub>-ATPases) that translocate PtdSer and other phospholipids from the luminal to the cytoplasmic leaflet<sup>32,33</sup>. Also of note is experimental evidence that PtdIns is not equally distributed over the ER but concentrated in highly dynamic sub-compartments<sup>34</sup>, suggesting that the cytoplasmic surface of the ER is overall relatively neutral. This is in sharp contrast to the cytoplasmic surface of the *trans*-Golgi and plasma membrane, at which PtdSer and phosphoinositides contribute to negative charge.

Thus, the Golgi defines a demarcation line between two broad membrane territories with distinct physical and functional features<sup>34</sup>. One is the territory of thin bilayers, loose lipid packing and neutral cytoplasmic surface charge, which is adapted for biogenic functions and comprises the ER and *cis*-Golgi. The other is the territory of thick bilayers, tight lipid packing and negative cytoplasmic surface charge, adapted for barrier functions and comprising the *trans*-Golgi and plasma membrane. The contrasting features of these territories provide specific cues for both integral and peripheral membrane proteins, allowing temporal and spatial control over the key events that occur along the secretory pathway. For instance, late secretory organelles recruit peripheral membrane proteins that contain polybasic motifs<sup>10,11</sup>. Early secretory organelles, on the other hand, bind proteins with neutral amphipathic lipid packing sensor (ALPS) motifs, containing bulky hydrophobic residues that readily insert into lipid-packing defects<sup>4,35</sup>. We discuss two basic mechanisms by which cells preserve the unique lipid compositions of early and late membrane territories. One depends on proteins that are capable of sensing alterations in lipid composition by monitoring physical properties of the membrane. The other uses pipelines to bypass vesicular connections for optimal control over lipid exchange between the two territories.

## Lipid composition sensors

The ER is the principal supplier of bulk lipids to other organelles. Its extensive network of tubular and planar membranes interfaces at contact sites with both secretory (for example, plasma membrane or *trans*-Golgi) and non-secretory organelles (for example, mitochondria or lipid droplets) to facilitate control over lipid transport and metabolism in response to varying cellular demands<sup>36,37</sup>. The ER is also exceptionally sensitive to perturbations in its unique lipid composition and biophysical properties. For instance, an imbalance between saturated and unsaturated phospholipids readily affects ER biogenic activity, inducing a stress response that can trigger cell death<sup>38,39</sup>. Interestingly, the biogenic membranes of some bacteria harbour thermosensors that control phospholipid desaturation. Of these, the thermosensor DesK from *Bacillus subtilis* is mechanistically the best understood. DesK seems to sense the thickening of the membrane as the temperature drops (Fig. 3a). This is thought to force a cluster of polar residues into a more apolar environment, resulting in a conformational change that triggers the autokinase activity of DesK<sup>40,41</sup>. Subsequently, phosphorylated DesK activates DesR, a transcriptional activator of a fatty acid desaturase. Promotion of membrane fluidity by the latter switches DesK from a kinase to a phosphatase, shutting down expression of the desaturase.

How the ER maintains an optimal level of phospholipid desaturation remains to be established, but there is increasing evidence that the underlying mechanisms are interconnected with the organelle's protein quality control system, which is called the unfolded protein response (UPR)<sup>42,43</sup>. The integral membrane protein IRE1 is an ER stress sensor that detects misfolded proteins in the ER lumen and activates UPR. This response activates transcription of lipid biosynthetic genes, allowing ER membrane expansion to counterbalance the protein-folding stress in its lumen<sup>44</sup>. Conversely, an excess of saturated fatty acids and other perturbations in phospholipid biosynthesis also activates the UPR<sup>39,43</sup>. Intriguingly, IRE1 that lacks the luminal

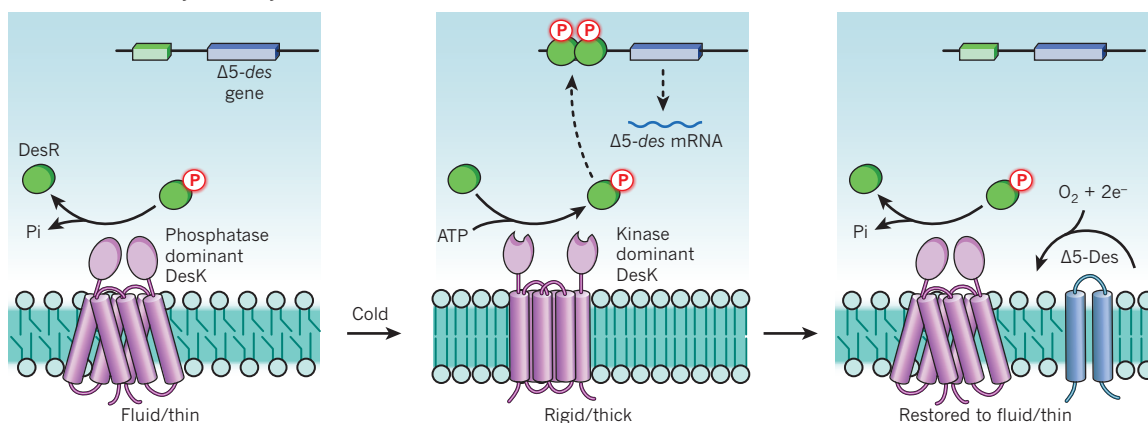
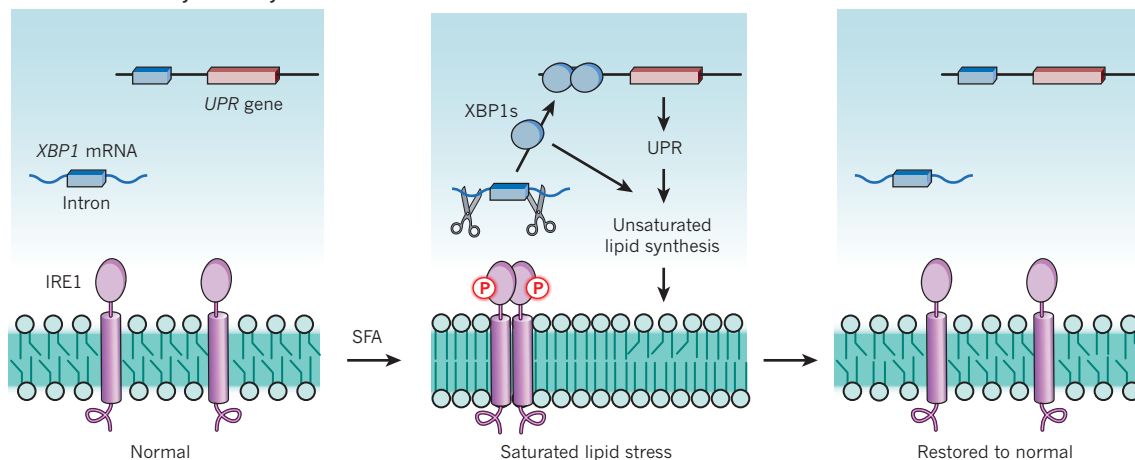
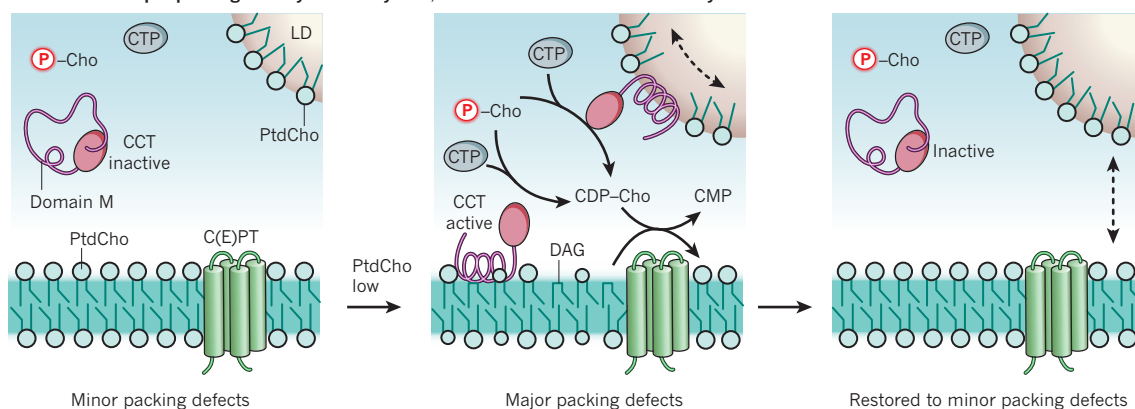
unfolded protein-sensing domain retains the ability to activate the UPR by lipid stress<sup>45,46</sup>, suggesting a dual role for IRE1 as membrane sensor of lipid perturbations in the ER (Fig. 3b).

Sterol excess blocks protein import into the ER<sup>47</sup>, whereas sterol depletion results in mistargeting of tail-anchored mitochondrial proteins to the ER<sup>48</sup>. Therefore, sterol levels in the ER of mammals are kept within narrow limits by a sophisticated feedback system that revolves around the membrane-bound transcription factor SREBP2 and the sterol-sensing protein SCAP. When ER sterol levels drop below 5 mol%, SREBP2 is escorted from the ER to the Golgi by SCAP<sup>21</sup>. In the Golgi, SREBP2 undergoes proteolytic cleavage to release a transcription factor that activates genes involved in sterol biosynthesis and uptake. Above the 5 mol% threshold, sterol-bound SCAP and SREBP2 remain in the ER, preventing SREBP2 processing. Thus, the SCAP–SREBP2-mediated feedback loop ensures a constant supply of ER-derived sterols to distal organelles within a range compatible with the biogenic function of the ER<sup>21</sup>. A second SREBP isoform, SREBP1, is part of an independent feedback loop that specifically responds to changes in PtdCho levels<sup>49</sup>. PtdCho depletion causes a redistribution of the SREBP-activating proteases from the Golgi to the ER, followed by proteolytic activation of SREBP1 and the induction of fatty-acid-biosynthetic genes. Apparently, the enhanced negative curvature stress associated with increasing the PtdEtn/PtdCho ratio affects curvature sensors that control COPI-mediated retrograde trafficking, causing a shift in the distribution of the SREBP-activating proteases towards the ER<sup>49</sup>.

Because PtdCho is a major membrane lipid, its production is a crucial point of control in global membrane homeostasis. The rate-limiting step in PtdCho biosynthesis is the formation of cytidine diphosphate-choline, a reaction catalysed by cytidine triphosphate (CTP):phosphocholine cytidylyltransferase (CCT; Fig. 2). CCT contains an amphipathic helix — called domain M — that silences the activity of the enzyme in its soluble form<sup>50</sup>. In the presence of PtdCho bilayers with a high content of conical lipids such as diacylglycerol or PtdEtn, domain M undergoes a conformational switch that creates a hydrophobic face for membrane binding<sup>51</sup>. Association with membranes relieves self-inhibition and drastically increases the affinity of CCT for its substrate CTP (Fig. 3c). By contrast, membrane association and activity of CCT is reduced when diacylglycerol/PtdCho or PtdEtn/PtdCho ratios are low. Thus, CCT seems to act as a general sensor of lipid packing defects that signals a demand for PtdCho biosynthesis to allow membrane expansion or to prevent its transition into a porous state. Consistent with this model, CCT is recruited to lipid droplets in the growing phase to allow production of more PtdCho molecules to surround the expanding oily core<sup>52</sup> (Fig. 3c).

## Lipid pipelines

Newly synthesized lipids are exported from the ER as components of secretory vesicles, or through pipelines operated by cytoplasmic lipid transfer proteins (LTPs)<sup>37,53</sup>. The latter mechanism is crucial for supplying ER lipids to mitochondria and other organelles that are not connected by vesicular trafficking but rely on lipid import for proper function<sup>54</sup>. Intriguingly, bulk transport of various lipids between the ER and plasma membrane continues unabated when vesicular trafficking is shut off<sup>55–58</sup>. Thus, lipid pipelines are also an important means by which the ER delivers lipids to distal secretory organelles. In fact, pipelines run from the ER to the *trans*-Golgi and plasma membrane to ferry sterols<sup>54,55,59</sup>, ceramide<sup>60</sup>, PtdCho<sup>57</sup>, PtdEtn<sup>58,61</sup>, PtdSer<sup>62</sup> and PtdIns<sup>63</sup>. A pipeline also transports glucosylceramide, the precursor of complex glycosphingolipids, to the *trans*-Golgi<sup>64,65</sup>. Pipelines require LTPs because spontaneous movement of most lipids between membranes occurs too slowly to support cell functions<sup>56,66,67</sup>. Beyond their obvious function in the supply of lipids, we propose that two other aspects of pipelines are key to membrane homeostasis. First, they ease lipid sorting and remodelling stress in the Golgi, thereby enabling early and late secretory organelles to maintain their

**a Membrane fluidity control by thermosensor DesK****b Membrane fluidity control by ER stress sensor Ire1****c Membrane lipid packing density control by CCT, which is needed for PtdCho biosynthesis****Figure 3 | Lipid composition sensors in biogenic organelles.**

**a**, Thermosensor DesK is a histidine kinase acting at the top of a regulatory cascade controlling the synthesis of unsaturated fatty acids in *Bacillus subtilis*. Following a drop in temperature, the membrane spans of the kinase sense an increase in bilayer thickness owing to an increased ordering of the lipid acyl chains. This promotes a kinase dominant state of DesK, which autophosphorylates and then transfers the phosphate group to DesR. Phospho-DesR activates transcription of the *des* gene coding for a  $\Delta 5$ -desaturase.  $\Delta 5$ -Des desaturates the acyl chains of membrane phospholipids, increasing membrane fluidity. This change promotes the phosphatase-dominant state of DesK, leading to DesR dephosphorylation and turning off transcription of the *des* gene. **b**, Ire1 is a conserved ER-resident stress sensor that can trigger an unfolded protein response (UPR) through a luminal domain that monitors the unfolded protein load in the ER. In addition, Ire1 uses its membrane span to detect membrane perturbations such as an increase in lipid saturation, which can signal UPR independently of the luminal stress-sensing domain. On activation, Ire1 is autophosphorylated and assembles into high-order oligomers.

Signalling proceeds by the non-conventional splicing of *XBP1* mRNA, which is initiated by Ire1's cytoplasmic RNase domain. Only the spliced mRNA is translated to produce an active transcription activator of UPR genes, which code for folding enzymes and lipid biosynthetic machinery, to restore ER membrane homeostasis. SFA, saturated fatty acid. **c**, CTP:phosphocholine cytidyltransferase (CCT), the rate-limiting enzyme in phosphatidylcholine (PtdCho) biosynthesis, interconverts between an inactive soluble and active membrane bound form in response to changes in membrane lipid packing density. An amphipathic helix, called domain M, silences the activity of the enzyme in its soluble form. Lipid packing defects arising from increased levels of conical lipids (for example, diacylglycerol (DAG), phosphatidylethanolamine (PtdEtn)) or by membrane expansion (for example, during lipid droplet (LD) growth) are sensed by domain M, creating a hydrophobic face for membrane binding. Association of domain M with membranes relieves self-inhibition, increasing the affinity of CCT for its substrate CTP. This then enhances biosynthesis of PtdCho to prevent transition of the membrane into a porous state. C(E)PT, choline/ethanolaminephosphotransferase.



compositional identity in the face of continuous vesicular trafficking. Second, they provide a rapid-response mechanism to facilitate regulatory crosstalk among the three main lipid classes. After discussing how pipelines operate, we consider how they could help to preserve the contrasting features of the early and late membrane territories of the secretory pathway.

At first approximation, LTPs are simply lipid exchangers, similar to cyclodextrins that bind and exchange hydrophobic compounds (Fig. 4a). Many LTPs were first identified by their ability to exchange lipids between vesicle populations in the test tube<sup>68</sup>. The exchange mechanism can be thought of as a multi-step process. Thus, a partially desorbed membrane lipid is reversibly taken up into the LTP's lipid binding pocket. A conformational change might then occur, causing the binding pocket to be closed off, protecting the lipid from the aqueous cytoplasm. The reverse process results in the release of lipid at an acceptor membrane. When extending this idea to cells, LTP-operated pipelines would function as bidirectional routes of lipid transfer, equilibrating lipids between the cytoplasmic faces of organelles. As already discussed, the situation is more complex because bulk lipid composition varies greatly along the secretory pathway, and transport often entails movement against a concentration gradient. Several factors must be considered in adapting this simple model of LTP-mediated lipid equilibration to the cellular situation.

First, many LTPs possess specific membrane interaction motifs such as the FFAT motif that engages ER receptors<sup>69</sup>, and a PH domain that binds phosphatidylinositol-4-phosphate (PtdIns(4)P) in late secretory organelles<sup>37</sup> (Fig. 4b). The presence of these motifs ensures a higher frequency of productive encounters between LTPs and particular membranes, ensuring specificity. In addition, LTPs, such as CERT<sup>60,70,71</sup> and oxysterol binding proteins (OSBPs)<sup>53</sup>, that have both motifs can engage two membranes simultaneously (Fig. 4c). These LTPs probably operate within the narrow (<30 nm) cytoplasmic gap created by membrane contact sites between the ER and *trans*-Golgi, potentially increasing the lipid exchange rate<sup>36</sup>. Second, transfer specificity alone cannot account for the fact that many lipids entering an ER pipeline eventually accumulate in an acceptor organelle. For this, the cell uses three general strategies: metabolic trapping, thermodynamic trapping and heterotypic lipid exchange. Ceramide transport provides a good example of metabolic trapping (Fig. 4c). CERT exchanges ceramide bidirectionally between the ER and *trans*-Golgi, but conversion of ceramide to sphingomyelin in the latter compartment ensures net one-way transport<sup>60</sup>. Thermodynamic trapping underlies a proposed model for sterol movement between the ER and the plasma membrane<sup>56</sup>. In this model, the association of sterols with saturated plasma membrane lipids (Fig. 1b) provides a thermodynamic trap. More precisely, the chemical activity of sterol in the ER and plasma membrane is the same, which is consistent with LTP-mediated equilibratory exchange of sterols between these compartments<sup>55,56</sup>, but the chemical activity coefficient for sterols in the plasma membrane is significantly lower than that in the ER, enabling higher plasma-membrane sterol concentrations<sup>72</sup>.

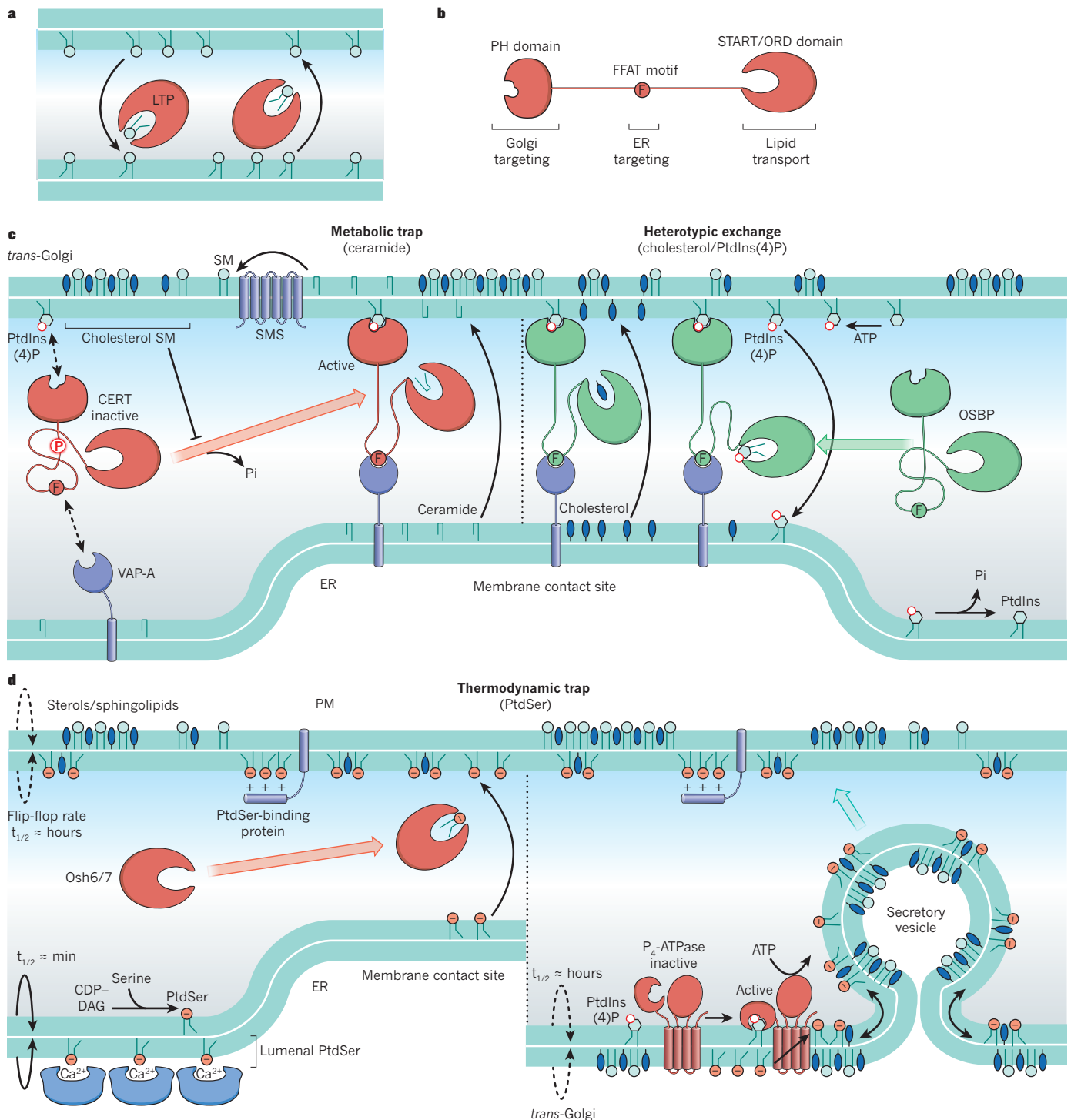
Thermodynamic traps might also help to explain how cells control the transbilayer and territorial asymmetry of PtdSer, as already discussed. Newly synthesized PtdSer is transported to the plasma membrane by secretory vesicles, as well as by a pipeline mechanism (Fig. 4d, right panel). PtdSer is carried in the luminal leaflet of secretory vesicles until it reaches the *trans*-Golgi at which it is flipped to the cytoplasmic leaflet by PtdIns(4)P-dependent P<sub>4</sub>-ATPases<sup>32,33</sup>. Golgi-derived vesicles then move PtdSer to the cytoplasmic face of the plasma membrane. Intersecting this delivery mechanism is a PtdSer pipeline operated by members of the OSBP family, such as Osh6 and Osh7 in yeast<sup>62</sup> (Fig. 4d, left panel). To maintain the cytoplasmically oriented pool of PtdSer in the plasma membrane in the presence of a PtdSer pipeline, it is likely that PtdSer is trapped by direct electrostatic interactions with the polybasic domains of plasma-membrane-localized proteins.

Heterotypic lipid exchange involves the reciprocal transfer of two different lipids by an LTP, as suggested for PtdIns–PtdCho exchangers<sup>37</sup>. Coupled with a metabolic trap for one of the lipids and a thermodynamic trap for the other, this exchange principle can account for net unidirectional movement of lipids, as recently shown for the transport of PtdIns(4)P and sterols by OSBP and its yeast counterpart Osh4 (refs 53, 73–75) (Fig. 4c). OSBP and Osh4 bind both sterol and PtdIns(4)P in a mutually exclusive manner. The proteins capture sterol in the ER and exchange it for PtdIns(4)P in the *trans*-Golgi. On returning to the ER, the reverse exchange occurs such that the LTP discharges PtdIns(4)P and is reloaded with sterol. The potential reversibility of this process is eliminated by a metabolic trap, which drives anterograde sterol movement. Thus, PtdIns(4)P is hydrolysed to PtdIns by an ER-localized phosphatase, ensuring that the LTP — which cannot bind PtdIns — departs the ER loaded with sterol. There are two further aspects of this model that must be considered. First, a thermodynamic trap for sterol at the *trans*-Golgi is necessary otherwise OSBP or Osh4 might not be able to acquire PtdIns(4)P efficiently at the *trans*-Golgi in the face of sterol excess. This requirement is satisfied by the local production of sphingolipids. Second, PtdIns pools must be regenerated at the *trans*-Golgi to enable fresh PtdIns(4)P synthesis. Because PtdIns is not synthesized in the *trans*-Golgi, it seems likely that the PtdIns generated by PtdIns(4)P hydrolysis in the ER is recycled to the *trans*-Golgi by a pipeline operated by PtdIns/PtdCho exchangers such as Sec14 or Nir2 (ref. 37).

## Regulatory crosstalk

The membrane contact site between the ER and *trans*-Golgi is a nexus for the coordinate regulation of sphingolipid and sterol levels that help to define late secretory organelles. The delivery of sterol to the *trans*-Golgi by OSBP in mammals and Osh4 in yeast activates a *trans*-Golgi localized PtdIns-4-kinase, causing an increase in PtdIns(4)P levels<sup>76</sup>. In mammals, this results in the recruitment of CERT (Fig. 4c) and promotes delivery of ceramide for sphingomyelin synthesis to set up a thermodynamic trap for sterols. A related process probably occurs in yeast, although the ceramide carrier has yet to be identified<sup>77</sup>. PtdCho is required in the PtdIns/PtdCho exchange process that restores Golgi PtdIns levels<sup>37</sup>, as well as for sphingomyelin biosynthesis (Fig. 2); because PtdCho is a bulk lipid, its high levels in the *trans*-Golgi ensure that these seemingly antagonistic demands are easily met. Thus, a number of pipelines intersect to ensure that sphingolipid precursors reach the *trans*-Golgi in synchrony with sterol arrival, allowing a fundamental transition in the lipid landscape that divides the secretory pathway in early and late membrane territories.

In addition to the coordinate regulation of the sphingolipid and sterol content of the plasma membrane, a remarkable homeostatic circuit regulates plasma membrane lipid organization by linking the transbilayer phospholipid asymmetry of the plasma membrane to its sphingolipid content<sup>78</sup> (Fig. 5). This circuit relies on a number of protein kinases, the action of which, in this context, has been best studied in yeast. Thus, the TORC2 protein kinase complex responds to plasma membrane sphingolipid levels: when sphingolipid levels are low, TORC2 phosphorylates the Ypk1 and Ypk2 kinases to generate two outcomes. The first is an increase in the production of sphingolipid precursors in the ER. The rate-limiting step in the synthesis of sphingolipid precursors is catalysed by serine palmitoyltransferase (SPT; Lcb1 or Lcb2 in yeast; Fig. 2) and regulated by Orm proteins<sup>79–81</sup>. SPT, Orm proteins and the PtdIns(4)P phosphatase Sac1 form the SPOTS complex in the ER membrane. SPT activity in the SPOTS complex is auto-inhibited, but this inhibition is relieved on Ypk1- and Ypk2-mediated phosphorylation of the Orm proteins. The second result of TORC2-mediated phosphorylation of Ypk1 and Ypk2 is to phosphorylate and downregulate the flippase kinases Fpk1 and Fpk2, which function as flippase activators<sup>82,83</sup>. Thus, sphingolipid levels in the exoplasmic leaflet of the plasma membrane are sensed and corrected by fresh synthesis of precursors. This occurs



**Figure 4 | Lipid pipelines operated by lipid transfer proteins.** **a**, Simple model for LTP-mediated lipid exchange, resulting in the equilibration of lipid cargo between donor and acceptor membranes. **b**, Schematic structure of LTPs, such as CERT and OSBP, that carry dual membrane targeting motifs as well as a lipid-binding domain. **c**, Ceramide and sterol pipelines at endoplasmic reticulum (ER)-*trans*-Golgi membrane contact sites in mammals. CERT (left) is inactive as a phospho-protein because of reciprocal masking of its PH domain and lipid-binding START domain<sup>96</sup>. On dephosphorylation, it simultaneously engages phosphatidylinositol 4-phosphate (PtdIns(4)P) at the *trans*-Golgi and VAP-A at the ER, catalysing movement of ceramide between the cytoplasmic faces of the two organelles. Ceramide flips to the luminal side of the *trans*-Golgi at which it is converted to sphingomyelin (SM) by its synthase (SMS). This metabolic trapping event enables continued delivery of ceramide from the ER. Like CERT, OSBP (right) exists in an inactive form in which its functional motifs are masked. Oxysterol binding relieves this autoinhibition enabling OSBP to

engage the ER-*trans*-Golgi contact site. OSBP uses heterotypic lipid exchange to move sterol between the ER and *trans*-Golgi, and PtdIns(4)P in the reverse direction. At the ER, PtdIns(4)P is hydrolysed to PtdIns. This is returned to the *trans*-Golgi by a PtdIns-phosphatidylcholine (PtdCho) exchanger (not shown), to be converted to PtdIns(4)P by a sterol-activated phosphatidylinositol-4-OH kinase (PI(4)K). **d**, A phosphatidylserine (PtdSer) pipeline operated by Osh6 and Osh7 in yeast transfers PtdSer synthesized from CDP-diacylglycerol (DAG) and serine on the cytoplasmic surface of the ER to the plasma membrane (PM) (left). Trapping events mediated by calcium-bound or positively charged proteins ensure that, at steady state, PtdSer is mainly localized in the lumen of the ER (despite the presence of a bidirectional flippase) and the cytoplasmic face of the plasma membrane. PtdSer in the luminal leaflet of the ER enters secretory vesicles and is delivered to the *trans*-Golgi (right). Flippase activity of P<sub>4</sub>-ATPases moves PtdSer to the cytoplasmic face. Thus, Golgi-derived transport vesicles carry PtdSer to the plasma-membrane cytoplasmic face.



in coordination with a  $P_4$ -ATPase-catalysed redistribution of aminophospholipids to the opposite leaflet, presumably to maximize impermeability of the plasma membrane (discussed later).

### System failure and disease

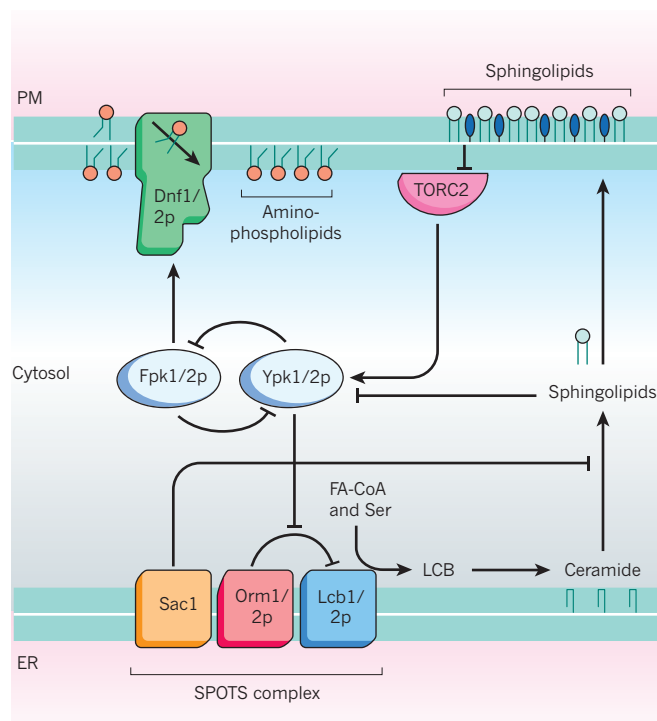
Numerous links between lipid imbalances and human pathologies underscore the importance of membrane lipid homeostasis<sup>1,84</sup>. We focus on two examples that highlight the physical principles of lipid organization in early and late membrane territories in the context of liver disease.

A surplus of free fatty acids contributes to liver failure in obesity and type 2 diabetes<sup>85,86</sup>. Although fatty-acid-induced liver disease is generally attributed to triglyceride accumulation in liver cells, recent data indicate that triglyceride accumulation might also have a protective role. The high toxicity of saturated fatty acids is partly due to a limited capacity of the liver cells to incorporate them into triglycerides<sup>86</sup>. Increased membrane saturation triggers ER stress (see earlier) and can profoundly disturb cellular homeostasis by affecting the function of membrane channels and transporters such as an ER-resident  $Ca^{2+}$ -pump, sarcoplasmic/endoplasmic reticulum calcium ATPase (SERCA)<sup>87,88</sup>. Inhibition of SERCA impairs protein folding in the ER and can trigger UPR. This, in turn, enhances SREBP1 expression and stimulates fatty acid synthesis<sup>89</sup>, creating a positive feedback loop that might further aggravate the metabolic derangements and liver injuries caused by saturated fatty acids. Thus, perturbations in ER lipid packing density and homeostasis can have far-reaching pathophysiological consequences. Further insight into the underlying molecular principles might reveal new opportunities for therapeutic interventions in common chronic diseases such as obesity and insulin resistance.

Alterations in the lipid packing density of late secretory organelles can have equally disastrous consequences. Progressive familial intrahepatic cholestasis-1 (PFIC1) is a potentially lethal liver disease caused by mutations in *ATP8B1* and characterized by a bile-salt secretion defect. *ATP8B1* codes for a  $P_4$ -ATPase that is thought to translocate PtdSer towards the cytoplasmic leaflet of the canalicular membrane of liver cells<sup>90</sup>. This membrane also harbours the bile salt export pump ABCB11. Interruption of the enterohepatic circulation of bile salts in patients with PFIC1 results in normalization of their hepatobiliary output, indicating that the bile-salt transport defect in PFIC1 is not a direct consequence of *ATP8B1* dysfunction<sup>91</sup>. The enhanced recovery of PtdSer in bile from *ATP8B1* mutant mice supports a role for *ATP8B1* as a PtdSer translocase<sup>90</sup>. These mice also have a drastic increase in biliary output of canalicular cholesterol. Subsequent studies showed that the activity of the bile-salt pump is crucially dependent on the cholesterol content of the canalicular membrane<sup>92</sup>. The exoplasmic leaflet of this membrane is rich in sphingolipids that are tightly packed with cholesterol to provide maximal resistance against the detergent action of hydrophobic bile salts. By flipping excess PtdSer from the exoplasmic surface towards the cytoplasmic leaflet, *ATP8B1* helps to preserve this barrier function. Loss of *ATP8B1* function results in lipid scrambling, thereby reducing the lipid ordering in the exoplasmic leaflet and increasing its sensitivity to hydrophobic bile salts. Increased cholesterol extraction by bile salts reduces the cholesterol content of the bilayer, which in turn impairs the activity of the bile salt pump and causes cholestasis<sup>90,92</sup>.

### Outlook

We have described the architecture of the secretory pathway in terms of contrasting membrane territories that are specialized for biogenic and barrier functions, and presented a homeostatic model in which lipid composition sensors and pipelines defend these territories against erosion by ongoing vesicular trafficking. Surprisingly, some recently identified lipid composition sensors have turned out to be dual function proteins, with their second function mediating seemingly unrelated processes such as protein quality control<sup>45,46</sup> and vesicular trafficking<sup>49</sup>. This reinforces the concept that membrane



**Figure 5 | Crosstalk between plasma membrane sphingolipid levels and transbilayer lipid asymmetry.** Low sphingolipid levels in the plasma membrane of yeast are sensed by the TORC2 protein kinase complex, which phosphorylates, and thus activates, the kinases Ypk1 and Ypk2. Ypk1/2-mediated phosphorylation events regulate sphingolipid synthesis and flippase activity. Synthesis of sphingolipid precursors is initiated by serine palmitoyltransferase (SPT; Lcb1 and Lcb2 in yeast), activity of which is inhibited in the SPOTS complex, which contains Orm1 and Orm2. Phosphorylation of Orm proteins by Ypk1/2 relieves SPT inhibition. Flippase activity is regulated by Fpk1- and Fpk2-mediated phosphorylation. Fpk1 and Fpk2 are downregulated on phosphorylation by Ypk1/2, resulting in a decrease in flippase activity. Although the SPOTS complex is conserved from mammals to yeast, Orm proteins in mammals lack the phosphorylation motifs involved in homeostatic regulation of sphingolipids. The Sac1 phosphatidylinositol 4-phosphate (PtdIns(4)P) phosphatase negatively regulates sphingolipid biosynthesis from ceramide, thus providing an additional level of homeostatic control. ER, endoplasmic reticulum; PM, plasma membrane.

lipid homeostasis is integral to a wide range of cellular processes. Because the dual roles of these proteins were not predicted, it seems likely that other examples of ‘moonlighting’ lipid composition sensors will be discovered.

Lipid pipelines coordinate lipid supply and demand between the territories of the early and late secretory pathway. The LTPs that operate these pipelines belong to large and poorly characterized protein families<sup>37</sup>, and occasionally are found to shuttle lipids other than those anticipated from their annotation<sup>62</sup>. Distinct from their role in transporting lipids, some LTP family members have been proposed to act as sensors of membrane lipid composition<sup>93</sup>. The ability of an LTP to extract lipid from a membrane could influence the nature of its interaction with downstream effectors, hence providing an indicator of membrane lipid status. Another interesting feature of some LTP family members is that they possess functional domains other than those required for lipid transfer. For example, the CERT-related proteins DLC1, DLC2 and DLC3 contain RhoGAP domains implicated in controlling cell migration and carcinogenesis<sup>94</sup>. The link between such dual functions is enigmatic and warrants further investigation.

There are numerous proposed intracellular lipid trafficking steps that are known to be non-vesicular, but for which LTPs have not yet been identified<sup>37</sup>. A prominent example of an ‘unassigned’ pipeline

is the one that moves sterols between the ER and plasma membrane. Although the reciprocal exchange of sterols and PtdIns(4)P by OSBP or Osh4 would result in the unidirectional movement of sterols from the ER to the *trans*-Golgi<sup>53,75</sup>, this mechanism cannot fully account for the equilibratory exchange of sterols between the ER and plasma membrane that is revealed by pulse-chase experiments<sup>55,56</sup>. Intriguingly, conditional elimination of the entire OSBP protein family in yeast has almost no impact on the exchange of sterols between the ER and plasma membrane<sup>95</sup>. Thus, the LTPs that operate the sterol pipeline between the ER and plasma membrane remain to be identified. The same holds for the lipid pipelines that run between the ER and non-secretory organelles, such as mitochondria, chloroplasts, lipid droplets and peroxisomes.

Further development of the concepts presented in this Review will require new assays to quantify lipid movement within cells. Although excellent *in vitro* studies have defined the main mechanistic steps of lipid transfer in quantitative detail, live cell tracking of lipids is fraught with complications. This certainly remains a key challenge for the future. ■

Received 2 February; accepted 3 April 2014.

- Wenk, M. R. The emerging field of lipidomics. *Nature Rev. Drug Discov.* **4**, 594–610 (2005).
- Shevchenko, A. & Simons, K. Lipidomics: coming to grips with lipid diversity. *Nature Rev. Mol. Cell Biol.* **11**, 593–598 (2010).
- Sharpe, H. J., Stevens, T. J. & Munro, S. A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* **142**, 158–169 (2010).  
**This comprehensive computational analysis reveals that membrane proteins show dichotomy in TMD length and composition that correlates with major differences in lipid composition and asymmetry between early and late secretory organelles.**
- Bigay, J. & Antonny, B. Curvature, lipid packing, and electrostatics of membrane organelles: defining cellular territories in determining specificity. *Dev. Cell* **23**, 886–895 (2012).
- Mesmin, B., Antonny, B. & Drin, G. Insights into the mechanisms of sterol transport between organelles. *Cell. Mol. Life Sci.* **70**, 3405–3421 (2013).
- de Kroon, A. I. P. M., Rijken, P. J. & De Smet, C. H. Checks and balances in membrane phospholipid class and acyl chain homeostasis, the yeast perspective. *Prog. Lipid Res.* **52**, 374–394 (2013).
- Koyanova, R. & Caffrey, M. Phases and phase transitions of the phosphatidylcholines. *Biochim. Biophys. Acta* **1376**, 91–145 (1998).
- Marsh, D. Lateral pressure profile, spontaneous curvature frustration, and the incorporation and conformation of proteins in membranes. *Biophys. J.* **93**, 3884–3899 (2007).
- Frolov, V. A., Shnyrova, A. V. & Zimmerberg, J. Lipid polymorphisms and membrane shape. *Cold Spring Harb. Perspect. Biol.* **3**, a004747 (2011).
- Magalhaes, M. A. O. & Glogauer, M. Pivotal advance: phospholipids determine net membrane surface charge resulting in differential localization of active Rac1 and Rac2. *J. Leukoc. Biol.* **87**, 545–555 (2010).
- Grinstein, S. Imaging signal transduction during phagocytosis: phospholipids, surface charge, and electrostatic interactions. *Am. J. Physiol. Cell Physiol.* **299**, C876–C881 (2010).
- van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nature Rev. Mol. Cell Biol.* **9**, 112–124 (2008).
- Slotte, J. P. Biological functions of sphingomyelins. *Prog. Lipid Res.* **52**, 424–437 (2013).
- Koyanova, R. & Caffrey, M. Phases and phase transitions of the sphingolipids. *Biochim. Biophys. Acta* **1255**, 213–236 (1995).
- Yeagle, P. *The Membranes of Cells* (Academic, 1993).
- Brown, D. A. & London, E. Structure and origin of ordered lipid domains in biological membranes. *J. Membr. Biol.* **164**, 103–114 (1998).
- Boumann, H. A. *et al.* Depletion of phosphatidylcholine in yeast induces shortening and increased saturation of the lipid acyl chains: evidence for regulation of intrinsic membrane curvature in a eukaryote. *Mol. Biol. Cell* **17**, 1006–1017 (2006).
- Hikiji, T., Miura, K., Kiyono, K., Shibuya, I. & Ohta, A. Disruption of the *CHO1* gene encoding phosphatidylserine synthase in *Saccharomyces cerevisiae*. *J. Biochem.* **104**, 894–900 (1988).
- Lester, R. L., Wells, G. B., Oxford, G. & Dickson, R. C. Mutant strains of *Saccharomyces cerevisiae* lacking sphingolipids synthesize novel inositol glycerophospholipids that mimic sphingolipid structures. *J. Biol. Chem.* **268**, 845–856 (1993).
- Park, E. & Rapoport, T. A. Mechanisms of Sec61/SecY-mediated protein translocation across membranes. *Annu. Rev. Biophys.* **41**, 21–40 (2012).
- Radhakrishnan, A., Goldstein, J. L., McDonald, J. G. & Brown, M. S. Switch-like control of SREBP-2 transport triggered by small changes in ER cholesterol: a delicate balance. *Cell Metab.* **8**, 512–521 (2008).  
**This study determined the narrow range within which ER cholesterol levels control SREBP-SCAP export from the ER through cooperative interactions between cholesterol, SCAP and INSIG.**

- Holthuis, J. C., Pomorski, T., Raggars, R. J., Sprong, H. & Van Meer, G. The organizing potential of sphingolipids in intracellular membrane transport. *Physiol. Rev.* **81**, 1689–1723 (2001).
- Schneider, R. *et al.* Electrospray ionization tandem mass spectrometry (ESI-MS/MS) analysis of the lipid molecular species composition of yeast subcellular membranes reveals acyl chain-based sorting/remodeling of distinct molecular species en route to the plasma membrane. *J. Cell Biol.* **146**, 741–754 (1999).
- Slotte, J. P. & Bierman, E. L. Depletion of plasma-membrane sphingomyelin rapidly alters the distribution of cholesterol between plasma membranes and intracellular cholesterol pools in cultured fibroblasts. *Biochem. J.* **250**, 653–658 (1988).
- Brügger, B. *et al.* Evidence for segregation of sphingomyelin and cholesterol during formation of COPI-coated vesicles. *J. Cell Biol.* **151**, 507–518 (2000).
- Munro, S. An investigation of the role of transmembrane domains in Golgi protein retention. *EMBO J.* **14**, 4695–4704 (1995).
- Quiroga, R., Trenchi, A., González Montoro, A., Valdez Taubas, J. & Maccioni, H. J. F. Short transmembrane domains with high-volume exoplasmic halves determine retention of Type II membrane proteins in the Golgi complex. *J. Cell Sci.* **126**, 5344–5349 (2013).
- Kaiser, H.-J. *et al.* Lateral sorting in model membranes by cholesterol-mediated hydrophobic matching. *Proc. Natl Acad. Sci. USA* **108**, 16628–16633 (2011).
- Fairn, G. D. *et al.* High-resolution mapping reveals topologically distinct cellular pools of phosphatidylserine. *J. Cell Biol.* **194**, 257–275 (2011).  
**This important study provides quantitative information on the transbilayer distribution of PtdSer in early and late secretory organelles with the aid of a genetically encoded lipid probe.**
- Sanyal, S. & Menon, A. K. Flipping lipids: why an' what's the reason for? *ACS Chem. Biol.* **4**, 895–909 (2009).
- Bartz, R. *et al.* Lipidomics reveals that adiposomes store ether lipids and mediate phospholipid traffic. *J. Lipid Res.* **48**, 837–847 (2007).
- Natarajan, P., Wang, J., Hua, Z. & Graham, T. R. Drs2p-coupled aminophospholipid translocase activity in yeast Golgi membranes and relationship to *in vivo* function. *Proc. Natl Acad. Sci. USA* **101**, 10614–10619 (2004).
- Alder-Baerens, N., Lisman, Q., Luong, L., Pomorski, T. & Holthuis, J. C. Loss of P<sub>4</sub> ATPases Drs2p and Dnf3p disrupts aminophospholipid transport and asymmetry in yeast post-Golgi secretory vesicles. *Mol. Biol. Cell* **17**, 1632–1642 (2006).
- Kim, Y. J., Guzman-Hernandez, M. L. & Balla, T. A highly dynamic ER-derived phosphatidylinositol-synthesizing organelle supplies phosphoinositides to cellular membranes. *Dev. Cell* **21**, 813–824 (2011).
- Vanni, S. *et al.* Amphipathic lipid packing sensor motifs: probing bilayer defects with hydrophobic residues. *Biophys. J.* **104**, 575–584 (2013).
- Levine, T. Short-range intracellular trafficking of small molecules across endoplasmic reticulum junctions. *Trends Cell Biol.* **14**, 483–490 (2004).
- Lev, S. Non-vesicular lipid transport by lipid-transfer proteins and beyond. *Nature Rev. Mol. Cell Biol.* **11**, 739–750 (2010).
- Diakogiannaki, E., Welters, H. J. & Morgan, N. G. Differential regulation of the endoplasmic reticulum stress response in pancreatic  $\beta$ -cells exposed to long-chain saturated and monounsaturated fatty acids. *J. Endocrinol.* **197**, 553–563 (2008).
- Deguil, J. *et al.* Modulation of lipid-induced ER stress by fatty acid shape. *Traffic* **12**, 349–362 (2011).
- Cybulski, L. E., Martin, M., Mansilla, M. C., Fernández, A. & de Mendoza, D. Membrane thickness cue for cold sensing in a bacterium. *Curr. Biol.* **20**, 1539–1544 (2010).
- Inda, M. E. *et al.* A lipid-mediated conformational switch modulates the thermosensing activity of DesK. *Proc. Natl Acad. Sci. USA* **111**, 3579–3584 (2014).  
**This elegant study revealed the molecular principles by which a membrane-bound thermosensor allows bacteria to adjust lipid desaturation for optimal membrane fluidity.**
- Jonikas, M. C. *et al.* Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**, 1693–1697 (2009).
- Surra, M. A. *et al.* A lipid E-MAP identifies Ubx2 as a critical regulator of lipid saturation and lipid bilayer stress. *Mol. Cell* **51**, 519–530 (2013).
- Schuck, S., Prinz, W. A., Thorn, K. S., Voss, C. & Walter, P. Membrane expansion alleviates endoplasmic reticulum stress independently of the unfolded protein response. *J. Cell Biol.* **187**, 525–536 (2009).
- Promlek, T. *et al.* Membrane aberrancy and unfolded proteins activate the endoplasmic reticulum stress sensor Ire1 in different ways. *Mol. Biol. Cell* **22**, 3520–3532 (2011).
- Volmer, R., van der Ploeg, K. & Ron, D. Membrane lipid saturation activates endoplasmic reticulum unfolded protein response transducers through their transmembrane domains. *Proc. Natl Acad. Sci. USA* **110**, 4628–4633 (2013).
- Nilsson, I., Ohvo-Rekilä, H., Slotte, J. P., Johnson, A. E. & von Heijne, G. Inhibition of protein translocation across the endoplasmic reticulum membrane by sterols. *J. Biol. Chem.* **276**, 41748–41754 (2001).
- Krumpe, K. *et al.* Ergosterol content specifies targeting of tail-anchored proteins to mitochondrial outer membranes. *Mol. Biol. Cell* **23**, 3927–3935 (2012).
- Walker, A. K. *et al.* A conserved SREBP-1/phosphatidylcholine feedback circuit regulates lipogenesis in metazoans. *Cell* **147**, 840–852 (2011).
- Lee, J., Taneva, S. G., Holland, B. W., Tieleman, D. P. & Cornell, R. B. Structural basis for autoinhibition of CTP:phosphocholine cytidyltransferase (CCT), the regulatory enzyme in phosphatidylcholine synthesis, by its membrane-binding amphipathic helix. *J. Biol. Chem.* **289**, 1742–1755 (2014).



51. Attard, G. S., Templer, R. H., Smith, W. S., Hunt, A. N. & Jackowski, S. Modulation of CTP:phosphocholine cytidylyltransferase by membrane curvature elastic stress. *Proc. Natl Acad. Sci. USA* **97**, 9032–9036 (2000).
52. Krahmer, N. *et al.* Phosphatidylcholine synthesis for lipid droplet expansion is mediated by localized activation of CTP:phosphocholine cytidylyltransferase. *Cell Metab.* **14**, 504–515 (2011).
53. Mesmin, B. *et al.* A four-step cycle driven by PI(4)P hydrolysis directs sterol/PI(4)P exchange by the ER-Golgi tether OSBP. *Cell* **155**, 830–843 (2013).
54. Tatsuta, T., Scharwey, M. & Langer, T. Mitochondrial lipid trafficking. *Trends Cell Biol.* **24**, 44–52 (2014).
55. Urbani, L. & Simoni, R. D. Cholesterol and vesicular stomatitis virus G protein take separate routes from the endoplasmic reticulum to the plasma membrane. *J. Biol. Chem.* **265**, 1919–1923 (1990).
56. Baumann, N. A. *et al.* Transport of newly synthesized sterol to the sterol-enriched plasma membrane occurs via nonvesicular equilibration. *Biochemistry* **44**, 5816–5826 (2005).  
**Together with ref. 55 this paper shows that anterograde sterol transport from the ER to the plasma membrane occurs by a non-vesicular, equilibratory mechanism in mammalian cells and yeast.**
57. Kaplan, M. R. & Simoni, R. D. Intracellular transport of phosphatidylcholine to the plasma membrane. *J. Cell Biol.* **101**, 441–445 (1985).
58. Sleight, R. G. & Pagano, R. E. Rapid appearance of newly synthesized phosphatidylethanolamine at the plasma membrane. *J. Biol. Chem.* **258**, 9050–9058 (1983).
59. Heino, S. *et al.* Dissecting the role of the golgi complex and lipid rafts in biosynthetic transport of cholesterol to the cell surface. *Proc. Natl Acad. Sci. USA* **97**, 8375–8380 (2000).
60. Hanada, K. *et al.* Molecular machinery for non-vesicular trafficking of ceramide. *Nature* **426**, 803–809 (2003).  
**This milestone paper defines the molecular mechanism for non-vesicular transport of ceramide from the ER to the trans-Golgi.**
61. Kobayashi, T. & Pagano, R. E. Lipid transport during mitosis. Alternative pathways for delivery of newly synthesized lipids to the cell surface. *J. Biol. Chem.* **264**, 5966–5973 (1989).
62. Maeda, K. *et al.* Interactome map uncovers phosphatidylserine transport by oxysterol-binding proteins. *Nature* **501**, 257–261 (2013).  
**This study shows that Osh6 and Osh7 transport PtdSer but not sterol between vesicle populations in the test tube and between the ER and plasma membrane in cells at ER-plasma-membrane contact sites.**
63. Schnabl, M., Daum, G. & Pichler, H. Multiple lipid transport pathways to the plasma membrane in yeast. *Biochim. Biophys. Acta* **1687**, 130–140 (2005).
64. Halter, D. *et al.* Pre- and post-Golgi translocation of glucosylceramide in glycosphingolipid synthesis. *J. Cell Biol.* **179**, 101–115 (2007).
65. D'Angelo, G. *et al.* Vesicular and non-vesicular transport feed distinct glycosylation pathways in the Golgi. *Nature* **501**, 116–120 (2013).
66. Silvius, J. R. & Leventis, R. Spontaneous interbilayer transfer of phospholipids: dependence on acyl chain composition. *Biochemistry* **32**, 13318–13326 (1993).
67. Estronca, L. M. B. B., Moreno, M. J. & Vaz, W. L. C. Kinetics and thermodynamics of the association of dehydroergosterol with lipid bilayer membranes. *Biophys. J.* **93**, 4244–4253 (2007).
68. Wirtz, K. W. Phospholipid transfer proteins. *Annu. Rev. Biochem.* **60**, 73–99 (1991).
69. Loewen, C. J. R., Roy, A. & Levine, T. P. A conserved ER targeting motif in three families of lipid binding proteins and in Opi1p binds VAP. *EMBO J.* **22**, 2025–2035 (2003).
70. Kawano, M., Kumagai, K., Nishijima, M. & Hanada, K. Efficient trafficking of ceramide from the endoplasmic reticulum to the Golgi apparatus requires a VAMP-associated protein-interacting FFAT motif of CERT. *J. Biol. Chem.* **281**, 30279–30288 (2006).
71. Kumagai, K., Kawano-Kawada, M. & Hanada, K. Phosphoregulation of the ceramide transport protein CERT at serine 315 in the interaction with VAMP-associated protein (VAP) for inter-organellar trafficking of ceramide in mammalian cells. *J. Biol. Chem.* **289**, 10748–10760 (2014).
72. Maxfield, F. R. & Menon, A. K. Intracellular sterol transport and distribution. *Curr. Opin. Cell Biol.* **18**, 379–385 (2006).
73. Ngo, M. & Ridgway, N. D. Oxysterol binding protein-related protein 9 (ORP9) is a cholesterol transfer protein that regulates Golgi structure and function. *Mol. Biol. Cell* **20**, 1388–1399 (2009).
74. Ridgway, N. D. Oxysterol-binding proteins. *Subcell. Biochem.* **51**, 159–182 (2010).
75. de Saint-Jean, M. *et al.* Osh4p exchanges sterols for phosphatidylinositol 4-phosphate between lipid bilayers. *J. Cell Biol.* **195**, 965–978 (2011).  
**Together with ref. 53, this study shows that a mammalian OSBP and yeast Osh4 bind PtdIns(4)P and sterols in a mutually exclusive manner to facilitate heterotypic lipid exchange between the ER and trans-Golgi.**
76. Banerji, S. *et al.* Oxysterol binding protein-dependent activation of sphingomyelin synthesis in the golgi apparatus requires phosphatidylinositol 4-kinase IIa. *Mol. Biol. Cell* **21**, 4141–4150 (2010).
77. Funato, K. & Riezman, H. Vesicular and nonvesicular transport of ceramide from ER to the Golgi apparatus in yeast. *J. Cell Biol.* **155**, 949–959 (2001).
78. Breslow, D. K. Sphingolipid homeostasis in the endoplasmic reticulum and beyond. *Cold Spring Harb. Perspect. Biol.* **5**, a013326 (2013).
79. Breslow, D. K. *et al.* Orm family proteins mediate sphingolipid homeostasis. *Nature* **463**, 1048–1053 (2010).
80. Han, S., Lone, M. A., Schneider, R. & Chang, A. Orm1 and Orm2 are conserved endoplasmic reticulum membrane proteins regulating lipid homeostasis and protein quality control. *Proc. Natl Acad. Sci. USA* **107**, 5851–5856 (2010).  
**Together with ref. 79, this study identified Orm proteins as homeostatic regulators of sphingolipid biosynthesis through formation of a complex with the rate-limiting enzyme SPT.**
81. Roelants, F. M., Breslow, D. K., Muir, A., Weissman, J. S. & Thorner, J. Protein kinase Ypk1 phosphorylates regulatory proteins Orm1 and Orm2 to control sphingolipid homeostasis in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **108**, 19222–19227 (2011).
82. Nakano, K., Yamamoto, T., Kishimoto, T., Noji, T. & Tanaka, K. Protein kinases Fpk1p and Fpk2p are novel regulators of phospholipid asymmetry. *Mol. Biol. Cell* **19**, 1783–1797 (2008).
83. Roelants, F. M., Baltz, A. G., Trott, A. E., Fereres, S. & Thorner, J. A protein kinase network regulates the function of aminophospholipid flippases. *Proc. Natl Acad. Sci. USA* **107**, 34–39 (2010).
84. Sanyal, A. J. Mechanisms of disease: pathogenesis of nonalcoholic fatty liver disease. *Nature Clin. Pract. Gastroenterol. Hepatol.* **2**, 46–53 (2005).
85. Kusminski, C. M., Shetty, S., Orci, L., Unger, R. H. & Scherer, P. E. Diabetes and apoptosis: lipotoxicity. *Apoptosis* **14**, 1484–1495 (2009).
86. Zámbo, V. *et al.* Lipotoxicity in the liver. *World J. Hepatol.* **5**, 550–557 (2013).
87. Li, Y. *et al.* Enrichment of endoplasmic reticulum with cholesterol inhibits sarcoplasmic-endoplasmic reticulum calcium ATPase-2b activity in parallel with increased order of membrane lipids: implications for depletion of endoplasmic reticulum calcium stores and apoptosis in cholesterol-loaded macrophages. *J. Biol. Chem.* **279**, 37030–37039 (2004).
88. Fu, S. *et al.* Aberrant lipid metabolism disrupts calcium homeostasis causing liver endoplasmic reticulum stress in obesity. *Nature* **473**, 528–531 (2011).
89. Fang, D.-L. *et al.* Endoplasmic reticulum stress leads to lipid accumulation through upregulation of SREBP-1c in normal hepatic and hepatoma cells. *Mol. Cell. Biochem.* **381**, 127–137 (2013).
90. Paulusma, C. C. *et al.* Atp8b1 deficiency in mice reduces resistance of the canalicular membrane to hydrophobic bile salts and impairs bile salt transport. *Hepatology* **44**, 195–204 (2006).
91. Kurbegov, A. C. *et al.* Biliary diversion for progressive familial intrahepatic cholestasis: improved liver morphology and bile acid profile. *Gastroenterology* **125**, 1227–1234 (2003).
92. Paulusma, C. C., de Waart, D. R., Kunne, C., Mok, K. S. & Elferink, R. P. J. O. Activity of the bile salt export pump (ABCB11) is critically dependent on canalicular membrane cholesterol content. *J. Biol. Chem.* **284**, 9947–9954 (2009).
93. Bankaitis, V. A. *et al.* Thoughts on Sec14-like nanoreactors and phosphoinositide signaling. *Adv. Biol. Regul.* **52**, 115–121 (2012).
94. Alpy, F. & Tomasetto, C. START ships lipids across interorganelle space. *Biochimie* **96**, 85–95 (2014).
95. Georgiev, A. G. *et al.* Osh proteins regulate membrane sterol organization but are not required for sterol movement between the ER and PM. *Traffic* **12**, 1341–1355 (2011).  
**This paper shows that the OSBP protein family in yeast is not required for sterol transport between the ER and plasma membrane.**
96. Saito, S. *et al.* Protein phosphatase 2C<sub>e</sub> is an endoplasmic reticulum integral membrane protein that dephosphorylates the ceramide transport protein CERT to enhance its association with organelle membranes. *J. Biol. Chem.* **283**, 6584–6593 (2008).

**Acknowledgements** Work in the authors laboratories is supported by the European Union (MC-ITN Sphingonet Project 289278 to J.C.H.), the Deutsche Forschungsgemeinschaft (SFB944-P14 to J.C.H.) and the National Institutes of Health (grant GM071041 to A.K.M.).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at [go.nature.com/zls5vn](http://go.nature.com/zls5vn). Correspondence should be addressed to J.C.H. ([holthuis@uos.de](mailto:holthuis@uos.de)) or A.K.M. ([akm2003@med.cornell.edu](mailto:akm2003@med.cornell.edu)).

# Sphingolipid metabolites in inflammatory disease

Michael Maceyka<sup>1</sup> & Sarah Spiegel<sup>1</sup>

**Sphingolipids are ubiquitous building blocks of eukaryotic cell membranes. Progress in our understanding of sphingolipid metabolism, state-of-the-art sphingolipidomic approaches and animal models have generated a large body of evidence demonstrating that sphingolipid metabolites, particularly ceramide and sphingosine-1-phosphate, are signalling molecules that regulate a diverse range of cellular processes that are important in immunity, inflammation and inflammatory disorders. Recent insights into the molecular mechanisms of action of sphingolipid metabolites and new perspectives on their roles in regulating chronic inflammation have been reported. The knowledge gained in this emerging field will aid in the development of new therapeutic options for inflammatory disorders.**

In the 1880s, the neurochemist J. L. W. Thudichum presciently named the brain lipid 'sphingosine' after the Sphinx, owing to its enigmatic chemical nature<sup>1</sup>. Sphingosine and its relatives continue to surprise and confound us today. These fatty amino alcohols are the backbone of a ubiquitous class of eukaryotic lipids, the sphingolipids, which include ceramide (*N*-acyl-sphingosine), sphingomyelin, and hundreds of different glycosphingolipids (Fig. 1a). Sphingolipids, like glycerolipids, are reservoirs of bioactive metabolites of profound importance to myriad cell signalling and pathological functions. The sphingolipid metabolites, ceramide, ceramide-1-phosphate (C1P), and sphingosine-1-phosphate (S1P) are emerging as important signalling molecules that regulate cell growth, survival, immune cell trafficking, and vascular and epithelial integrity, and are particularly important in inflammation and cancer<sup>2,3</sup>.

The past decade has seen an explosive advancement in the field of sphingolipid signalling based on the convergence of several key aspects. First, most of the regulatory proteins and enzymes involved in sphingolipid metabolism and the receptors for S1P have been cloned. This allowed the generation of knockout mice, yielding insights into the physiological functions of sphingolipid metabolites. Second, the advent of advanced mass spectroscopic techniques has brought the 'omics' revolution to sphingolipids, allowing for the simultaneous analysis and quantification of multiple species. Third, specific agonists and antagonists of S1P receptors and inhibitors of signalling enzymes were developed. The chief development among these was the discovery of FTY720 (fingolimod), a sphingosine analogue that alters immune cell trafficking and is already being used in the clinic for the treatment of multiple sclerosis<sup>4</sup>.

These are exciting times for the field and research continues apace. Several sphingolipid signalling protein structures have been solved, allowing for rational drug design. This Review will focus on the function of three key bioactive sphingolipids: ceramide, C1P and S1P, and their roles in inflammation. Although this is a normal physiological response to harmful stimuli such as infection, unchecked inflammation can lead to numerous pathophysiological states, including oedema, asthma, inflammatory bowel disease and associated cancer, and autoimmune disorders such as multiple sclerosis and rheumatoid arthritis. Sphingolipid metabolites play crucial parts at multiple stages of these disorders, and new mechanistic

perspectives on their actions will be discussed. We will also highlight how knowledge gained in this relatively new field will aid in the development of therapeutic options for inflammatory disorders.

## Sphingolipid metabolism

Sphingolipids are essential lipids consisting of a sphingoid backbone that is *N*-acylated with various fatty acids to form many ceramide species, which can have hundreds of distinct head groups. As this Review is mainly concerned with sphingolipid signalling and inflammatory diseases, we will focus on mammalian sphingolipids and their metabolism. Sphingolipid synthesis starts in the endoplasmic reticulum (ER) with the condensation of serine and palmitoyl coenzyme A (CoA) by serine palmitoyltransferase (SPT), the rate-limiting step, forming 3-ketosphinganine (Fig. 1b). SPT mutations can shift substrate preference to alanine, producing neurotoxic deoxysphingolipids that cause hereditary sensory and autonomic neuropathy type 1 (ref. 5). SPT activity is negatively regulated by ORMDL proteins<sup>6</sup>, and a genetic variant that increases ORMDL3 expression has been linked to increased risk of asthma<sup>7</sup>. Reduction of 3-ketosphinganine forms dihydrosphingosine that is then *N*-acylated by one of six ceramide synthases (CerS1–CerS6), each using specific acyl chains, typically with saturated or mono-unsaturated fatty acids with 14 to 26 carbons, forming dihydroceramides that are subsequently dehydrogenated to ceramides by dihydroceramide desaturase. Metabolism of ceramide to sphingomyelin and a vast array of complex glycosphingolipids mainly occurs in the Golgi (Fig. 2). ER to Golgi transport of ceramide is mediated by ceramide transfer protein (CERT) for sphingomyelin synthesis<sup>8</sup>, or by vesicular transport for glucosylceramide synthesis. Ceramides can also be phosphorylated in the Golgi by ceramide kinase (CERK) to form C1P, a rare species. Glycosphingolipid synthesis requires the transfer of glucosylceramide to the *trans*-Golgi network (TGN) by four-phosphate adaptor protein 2 (FAPP2), which also regulates vesicular trafficking from the Golgi to the plasma membrane<sup>9</sup>.

Sphingolipids have a rapid turnover and their levels are controlled by the balance between synthesis and degradation in multiple compartments<sup>2</sup>. Sphingolipids are degraded in lysosomes by glycosidases or acid sphingomyelinases that remove the head groups to form ceramides. Deacylation of ceramide by ceramidases is the only pathway known to generate sphingosine. Sphingosine can be recycled back to ceramide, and 50% or more of sphingosine molecules follow this reutilization pathway, which has a significant role in sphingolipid homeostasis<sup>2</sup>. Sphingosine

<sup>1</sup>Department of Biochemistry and Molecular Biology, Virginia Commonwealth University School of Medicine, Richmond, Virginia 23298, USA.



can also be phosphorylated by one of two sphingosine kinases, SphK1 and SphK2, forming S1P, a pleiotropic bioactive metabolite and key intermediate in the sphingolipid-to-glycerolipid metabolic pathway<sup>10</sup>.

S1P can then either be dephosphorylated by phosphatases, including two S1P-specific, ER-localized phosphatases (SPP1 and SPP2), or irreversibly degraded by S1P lyase (SPL) to phosphoethanolamine and hexadecenal that can be incorporated into glycerolipids. The fatty aldehyde dehydrogenase ALDH3A2 converts hexadecenal to hexadecenoate, which is utilized for the formation of palmitoyl-CoA in glycerolipid synthesis, revealing the connection between sphingolipid and glycerolipid homeostasis<sup>11</sup>. ALDH3A2 is non-functional in Sjögren–Larsson syndrome, a disease characterized by ichthyosis and intellectual disability, suggesting that the accumulation of S1P degradation products contributes to the pathogenesis<sup>11</sup>.

## The S1P gradient

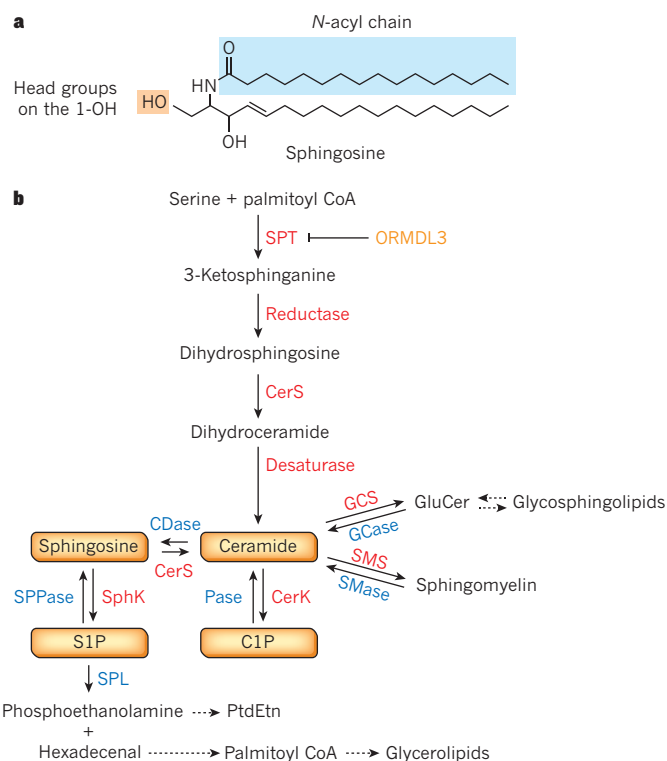
S1P concentrations are high in blood and lymph, and low in tissues, presumably due to the higher activity of S1P degrading enzymes in tissue. The high level of blood S1P has important homeostatic functions in the maintenance of vascular integrity, and the S1P gradient is crucial for immune cell trafficking. S1P in blood is rapidly turned over<sup>12</sup>, suggesting robust mechanisms for its synthesis inside cells and its transport outside. Both SphK1 and SphK2 contribute to circulating pools of S1P. Erythrocytes<sup>13</sup> and vascular endothelial cells<sup>14</sup> are the main sources of blood S1P, and lymphatic endothelial cells are the main sources for lymph S1P<sup>15</sup>. S1P is exported out of vascular and lymphatic endothelial cells by the specific transporter Spns2, which is not present in platelets or erythrocytes, suggesting the existence of other transporters in these cells<sup>16,17</sup>. Although ceramide, which is also abundant in plasma<sup>18</sup>, is associated with very-low-density lipoprotein and low-density lipoprotein, S1P in plasma is bound to both albumin and apolipoprotein M (apoM), which preferentially associates with high-density lipoprotein. However, only high-density lipoprotein containing apoM–S1P is required to preserve endothelial barrier integrity<sup>19</sup>. ApoM produced by the liver influences plasma S1P levels by enhancing S1P biosynthesis for secretion by hepatocytes, demonstrating that the liver is involved in S1P dynamics and that apoM delivers S1P to extrahepatic tissues<sup>20,21</sup>.

## Sphingolipid signalling

Multiple stimuli affect sphingolipid biosynthesis and metabolism by regulating key enzymes in a spatial and temporal manner, leading to the restricted production of ceramide, C1P and S1P<sup>2</sup>. For example, pathogens, oxidative stress and cytokines activate acid or neutral sphingomyelinases to generate ceramide in specific compartments<sup>2</sup>. Sphingosine generated by ceramidase is another sphingolipid metabolite involved in cell signalling<sup>22</sup> (see Review by Platt on page 68). However, its role in inflammatory responses is not yet clear<sup>23,24</sup>. Some inflammatory mediators also activate SphK1 by extracellular signal regulated kinase (ERK)-mediated phosphorylation, which promotes SphK1's translocation to the plasma membrane, at which its substrate, sphingosine, is localized and/or generated, resulting in transient elevations in S1P levels<sup>3,25</sup>. Although these studies were mainly carried out with cultured cells, the development of high-sensitivity mass spectrometry methods has led to demonstrations of changes in levels of bioactive sphingolipids being reported in specific tissues from mouse inflammatory disease models<sup>26–29</sup> and, in a few cases, in human samples<sup>30,31</sup>. More studies are needed to rigorously correlate bioactive lipids with disease stage. Understanding what bioactive sphingolipids do and how they signal is crucial for our understanding of their functions in inflammatory responses.

## Ceramide

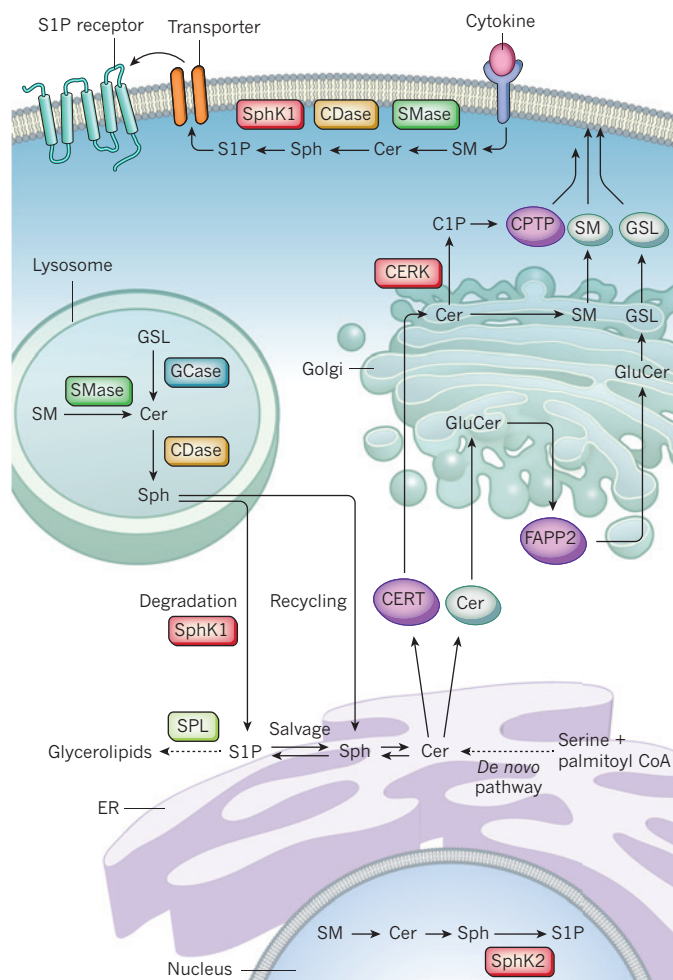
Despite the involvement of ceramide in numerous biological processes, only a few of its direct targets have been described. The best examples are ceramide-activated serine/threonine protein



**Figure 1 | Sphingolipid metabolism and interconnected bioactive sphingolipid metabolites.** **a**, Structure of ceramide, with a sphingosine backbone. **b**, *De novo* sphingolipid biosynthesis starts with the condensation of palmitoyl coenzyme A (CoA) and serine by serine palmitoyltransferase (SPT), an enzyme that is negatively regulated by ORM1-like protein 3 (ORMDL3). This is followed by a series of reactions (catalysed by enzymes in red) leading to formation of ceramide and subsequent formation of sphingomyelin and glycosphingolipids. Ceramide can be metabolized to other bioactive sphingolipid species, phosphorylated by ceramide kinase (CERK) to ceramide-1-phosphate (C1P), or hydrolysed to sphingosine, which is then phosphorylated to sphingosine-1-phosphate (S1P) by sphingosine kinases (SphKs). S1P can be degraded by phosphatases to sphingosine or by the lyase (SPL) that cleaves it to phosphoethanolamine and hexadecenal, which are subsequently reincorporated into glycerolipid metabolic pathways. For simplicity, degradative enzymes (blue) for reutilization of sphingolipids in the salvage pathway are included but these reactions take place in different subcellular compartments (see Fig. 2). CDase, ceramidase; CerS, ceramide synthase; GCse, glucosylceramidase; GCS, glucosylceramide synthase; Pase, phosphatase; PtdEtn, phosphatidylethanolamine; SMase, sphingomyelinase; SMS, sphingomyelin synthase; SPPase, sphingosine phosphate phosphatase.

phosphatases, including PP1, PP2A and PP2C<sup>32,33</sup>. PP2A is found in complex with SET (also known as I2PP2A), which inhibits PP2A function. Ceramide binds to SET, relieving its inhibitory actions<sup>33</sup>. Activation of PP2A that leads to dephosphorylation of Akt, a potent promoter of cell survival, could partly explain the pro-apoptotic actions of ceramide. In addition, PP1 could be the target of ceramide generated at the plasma membrane<sup>32</sup>. Stimulation of the atypical protein kinase PKC $\zeta$  by ceramide, probably concomitant with its membrane recruitment, also leads to inhibition of Akt<sup>34</sup>.

Another concept that has been subject to renewed interest is the formation of ceramide-enriched membrane microdomains or platforms as a mechanism by which ceramide transduces intracellular signalling. Alterations in membrane domains, owing to the elevation of ceramide and its tendency to self-associate, influence membrane composition and interactions of lipids or signalling proteins. These ceramide-rich platforms have been implicated in a variety of signalling cascades in immune cells, including activation of B cells, bacterial pathogen infections and release of cytokines during infection; they are also especially important in the induction of apoptosis<sup>35</sup>.



**Figure 2 | Subcellular compartmentalization of sphingolipid metabolism.** *De novo* ceramide (Cer) synthesis takes place in the endoplasmic reticulum (ER). Cer is delivered by ceramide transport protein (CERT) or vesicular transport to the Golgi for synthesis of ceramide-1-phosphate (C1P) (by ceramide kinase, CERK), sphingomyelin (SM), and glucosylceramide (GluCer). Four-phosphate adaptor protein 2 (FAPP2) then transports GluCer to the *trans*-Golgi for biosynthesis of complex glycosphingolipids (GSLs). SM and GSLs are delivered to the plasma membrane by vesicular transport and C1P by a C1P-specific transfer protein (CPTP). For signalling at the plasma membrane, sphingomyelinase (SMase), ceramidase (CDase), and sphingosine kinases (SphK) produce the bioactive metabolites Cer, sphingosine (Sph) and sphingosine-1-phosphate (S1P), respectively. S1P is then transported across the membrane. Membrane sphingolipids are internalized by the endocytic pathway and in the lysosome they are degraded by acidic forms of SMase, glycosidase (GCase) and CDase. The Sph formed can be metabolized to glycerolipids after phosphorylation by SphKs (probably SphK1) and cleavage by S1P lyase (SPL) or reutilized for sphingolipid synthesis in the salvage pathway. In the nucleus, SphK2-produced S1P inhibits histone deacetylases.

## C1P

C1P is formed on the TGN by CERK. Here, C1P activates cytosolic phospholipase-A<sub>2</sub> (cPLA<sub>2</sub>)<sup>36</sup>, the enzyme that releases the eicosanoid precursor arachidonate. Although early concerns regarding cPLA<sub>2</sub> activation by C1P were raised<sup>37</sup>, others have shown that not only does C1P bind cPLA<sub>2</sub> through an RxRH motif in its amino-terminal CaLB (also known as C2 lipid-binding domain) and recruit it to the membrane at which its phospholipid substrates reside, C1P also stimulates its activity and therefore has been implicated in eicosanoid production<sup>36,38</sup>. Conversely, C1P has been shown to inhibit ADAM17 (also known as tumour necrosis factor (TNF)-converting enzyme, or TACE), the major metalloprotease responsible for

cleaving pro-TNF to release the active inflammatory form, probably by binding to one or more similar tribasic motifs<sup>39</sup>.

## S1P

The many diverse roles of S1P in innate and adaptive immunity, including immunosurveillance, immune cell trafficking and differentiation, immune responses and endothelial barrier integrity are mediated by its binding to one of five G-protein-coupled receptors, named S1PR1–S1PR5 (refs 3, 40, 41). Downstream signalling of these receptors is complex as they are differentially expressed in immune and endothelial cells and couple to a varied set of heterotrimeric G proteins. For example, activation of S1PR1 generally promotes cell migration and the egress of T and B lymphocytes from lymphoid tissues<sup>42</sup>, whereas S1PR2 inhibits motility to promote retention of B cells in germinal centres<sup>43</sup>. S1P produced inside cells by activation of SphK1 can be secreted by SPNS2 or promiscuous ABC transporters<sup>46,44</sup>. S1P, in turn, signals through its receptors in a paracrine and/or autocrine manner, termed ‘inside-out’ signalling. However, the crystal structure of S1PR1 suggests that S1P must slide laterally within the plane of the bilayer to access the binding pocket<sup>45</sup>, raising the question of whether S1P needs to be secreted for its autocrine actions.

Much less is known about the intracellular targets of S1P. Within the past decade, new intracellular targets of S1P have been characterized. S1P formed by SphK1 in response to TNF or interleukin-1 (IL-1) binds to TNF receptor-associated factor 2 (TRAF2) and cellular inhibitor of apoptosis 2 (cIAP2), respectively, and enhances their lysine-63-linked polyubiquitylation activities<sup>46,47</sup>. In response to IL-1, which has a pivotal role in autoinflammatory diseases, cIAP2 and SphK1 form a complex with interferon regulatory factor-1 (IRF1), leading to its polyubiquitylation and activation. Consequently, IRF1 enhances expression of the chemokines CXCL10 and CCL5 that recruit mononuclear cells to sites of sterile inflammation<sup>47</sup>. Thus, in addition to canonical S1P receptor signalling, these new intracellular functions of S1P might be involved in the immune and inflammatory responses of these potent cytokines. Furthermore, S1P in the nucleus, formed by SphK2 or by inhibition of SPL, binds and inhibits the histone deacetylases HDAC1 and HDAC2, linking sphingolipid metabolism to inflammatory and metabolic gene expression<sup>48–50</sup>. Further studies are needed to solidify these new concepts.

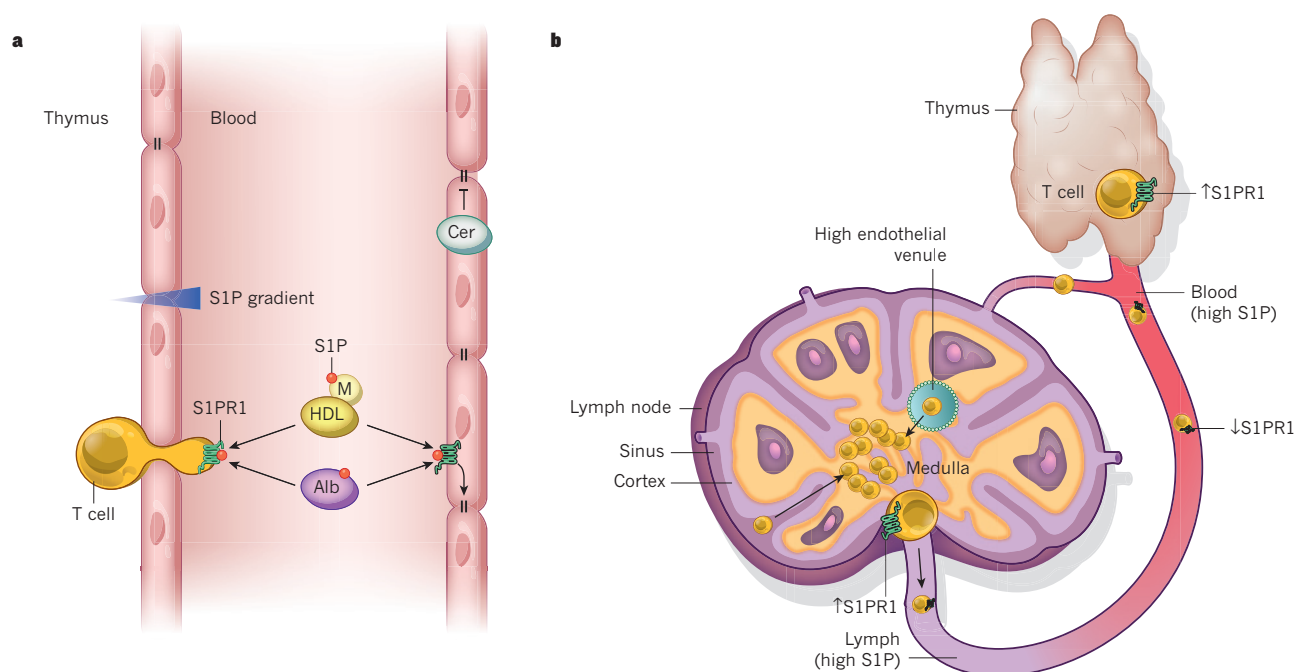
## Sphingolipids in endothelial biology and function

The vascular endothelium forms a barrier that lines blood vessels and plays an important part in cardiovascular homeostasis and blood flow, and in controlling the passage of leukocytes into and out of the bloodstream. Ceramide and S1P have opposing effects on vascular endothelium functions: whereas increased ceramide leads to endothelial barrier dysfunction<sup>51</sup>, S1P is crucial for maintenance of vascular integrity<sup>52–54</sup> (Fig. 3a).

Pro-inflammatory cytokines, oxidative stress and injurious insults, such as platelet activating factor (PAF), TNF and lipopolysaccharide, generate excess ceramide in endothelial cells linked to increased vascular permeability and dysfunction (Fig. 3a). Generation of ceramide is both stimulus- and context-specific, and is due to the activation of acid or neutral sphingomyelinase, or increased *de novo* biosynthesis. Ceramide triggers several pathways that induce endothelial cell death, including activation of caspases, or PP1 or PP2A<sup>2,55</sup>, and increasing mitochondrial permeability by forming ceramide-enriched platforms capable of translocating proteins. Moreover, PAF-induced formation of ceramide microdomains drives endothelial nitric oxide synthase (eNOS) activation and contributes to barrier dysfunction<sup>56</sup>. Ceramides have also been linked to growth arrest, cytoskeleton rearrangements, oxidative stress and senescence of endothelial cells<sup>2</sup>. Thus, ceramides regulate important endothelial cell functions that are thought to be responsible for the pathogenesis associated with vascular dysfunctions, including emphysema, sepsis and acute respiratory distress syndrome.

Using animal models of acute and chronic inflammation, it has been convincingly demonstrated that plasma S1P limits disruption of vascular endothelial monolayers and reduces oedema<sup>57</sup> (Fig. 3a).





**Figure 3 | S1P and S1PR1 control lymphocyte trafficking and vascular integrity.** **a**, The sphingosine-1-phosphate (S1P) level in lymphoid tissues, such as the thymus, is low compared with the blood, forming an S1P gradient (shaded red) that attracts lymphocytes and promotes S1PR1-dependent egress into the blood. In the blood, S1P is bound to albumin (Alb) and apolipoprotein M (M) in high-density lipoprotein (HDL), and signalling by S1PR1 maintains endothelium barrier function by promoting cell–cell interactions (right). Conversely, ceramide (Cer) can promote barrier leakage. **b**, When T cells are

ready to exit the thymus and enter the blood, S1PR1 is re-expressed so that T cells can respond to the chemotactic effect of high S1P levels in circulation. When T cells re-enter the blood, S1P downregulates S1PR1. Signals that control lymphocyte entry through high endothelial venules and retention in lymph nodes are not shown. Lymph node egress into lymphatic vessels also requires S1PR1. Increased lymphoid tissue S1P, by SPL inhibition, or in the presence of S1PR1 modulators such as FTY720, blocks egress of lymphocytes by disruption of the S1P gradient or desensitization of S1PR1 on T cells, respectively.

S1P activates endothelial S1PR1, leading to enhanced Rac-dependent cytoskeleton rearrangements, contacts between cells and the matrix, adherens junction assembly and barrier integrity<sup>3,41</sup>.

Lymphocytes circulate through lymph nodes for immune surveillance, entering at high endothelial venules (HEVs) — specialized blood vessels. Until recently, it was not known how HEVs allow lymphocyte transmigration, which increases during immune responses, and maintain vascular integrity. A study demonstrated that podoplanin expressed on HEV fibroblastic reticular cells binds and activates platelet C-type lectin-like receptor-2 (CLEC2)<sup>58</sup>. Activation of CLEC2 on extravasated platelets leads to the release of S1P in the perivascular space of HEVs. S1P, in turn, enhances vascular endothelial (VE)-cadherin expression for maintenance of the integrity of HEVs during immune responses<sup>58</sup>. It is likely that the S1P–S1PR1 axis is also involved in regulating the integrity of HEVs.

During inflammation, infiltrating immune cells and the production of pro-inflammatory cytokines increase endothelial permeability. S1P signalling attenuates the increased permeability induced by the inflammatory mediators histamine and PAF<sup>57</sup>. It is still not clear whether levels of S1P in plasma are reduced in systemic inflammation or sepsis. However, because levels of the S1P carrier protein apoM are decreased in patients with sepsis and systemic inflammatory response syndrome<sup>59</sup>, it is possible that S1P plasma levels are reduced in these conditions and thus contribute to vascular leakage.

Although S1PR2 does not regulate basal permeability, it protects against acute vascular barrier disruption and lethality after antigen challenge or PAF administration<sup>60</sup>. Endothelial S1PR2 inhibits Akt activation, suppressing eNOS and NO generation, and protects against the disassembly of adherens junctions by decreasing S-nitrosylation of  $\beta$ -catenin and thereby vascular leakage<sup>41,60</sup>. By contrast, during endotoxaemia, endothelial S1PR2 enhances vascular permeability and expression of pro-inflammatory and pro-coagulation molecules by activating Rho and p38 SAPK signalling pathways<sup>61</sup>. Together, these observations

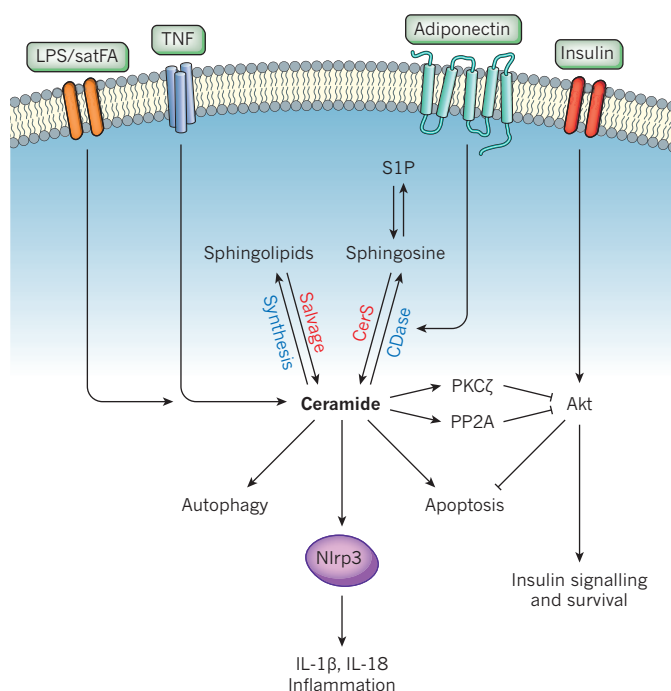
suggest that S1PR1 and S1PR2 have specific roles in preventing or triggering vascular leakage, probably by activation of distinct signalling pathways. For example, S1PR1 mediates G<sub>i</sub>-dependent Rac and Akt activation, whereas S1PR2 inhibits them and stimulates Rho<sup>3,41</sup>. Alternatively, the S1P receptors might be temporally and spatially localized and upregulated during inflammation.

An unexpected function of endothelial S1PR1 was discovered using S1PR1-specific agonists that suppressed the cytokine storm and recruitment of innate immune cells to the lung during influenza infection<sup>62</sup>. The efficacy of S1PR1 agonists was a result of the inhibition of common pathways downstream of multiple innate sensing pathways that are important for interferon production<sup>63</sup>. A better understanding of how S1P receptors and their downstream signalling are regulated and how they affect the vascular system could provide the basis for developing improved therapies for vascular disorders.

### Sphingolipids and inflammation

During inflammation, innate and adaptive immune cells enter sites of infection or injury and the activation of cytokine networks that follows helps to protect the host. Sphingolipid metabolites have key roles in the regulation of both trafficking and functions of immune cells<sup>42,64</sup>, and there are indications that sphingolipid metabolism and S1P receptors might be altered by inflammation.

S1P drives the differentiation of many types of immune cells, inducing changes in their functional phenotypes and regulating production of pro-inflammatory cytokines and eicosanoids. These topics will not be covered here as they have recently been extensively discussed (see refs 3, 42, 64 and 65 for authoritative reviews). In particular, S1P has emerged as a central regulator of lymphocyte egress. Although the S1P–S1PR1 axis regulates trafficking and migration of most types of immune cells (such as T and B lymphocytes, haematopoietic progenitors, macrophages, dendritic cells, neutrophils, mast cells, natural killer T cells and osteoclasts<sup>3,42</sup>), the regulation



**Figure 4 | Ceramide at the nexus of obesity and inflammation.** Ceramide is generated in response to obesity signals (such as saturated fatty acids, satFA), lipopolysaccharide (LPS) or pro-inflammatory cytokines (such as TNF) by enhancement of *de novo* biosynthesis or sphingolipid recycling. Ceramide can then signal to a variety of pathways that promote inflammation and the deleterious effects of obesity. Ceramide can bind to and inhibit SET/I2PP2A (not shown), leading to the activation of PP2A, which dephosphorylates and inhibits Akt, blocking many of the actions of insulin signalling and leading to insulin resistance, and decreased cell survival; activate PKC $\zeta$ , also leading to the inhibition of Akt; induce autophagy, which can influence innate immune responses; activate the Nlrp3 inflammasome to promote the processing and secretion of the pro-inflammatory cytokines IL-1 $\beta$  and IL-18; and induce apoptosis. Binding of adiponectin to its receptors stimulates ceramidase (CDase), degrading ceramide to sphingosine, which is then phosphorylated to sphingosine-1-phosphate (S1P). Insulin stimulates Akt. It should be noted that this diagram is not meant to imply that all of these events occur in the same cell. CerS, ceramide synthase.

of T- and B-lymphocyte trafficking is the best understood and will be used to illustrate the primary mechanisms involved.

### S1P and lymphocyte trafficking

S1P concentrations in blood and lymph are much greater than in tissues, and lymphocyte egress from the thymus or secondary lymphoid tissues is dependent on the cells sensing this gradient through activating S1PR1 (ref. 42). When ready to leave the thymus, T lymphocytes upregulate S1PR1 (ref. 66) and exit in response to the S1P gradient (Fig. 3b). S1PR1-dependent Rac activation is required for *trans*-endothelial migration of T cells and lymph node egress<sup>67</sup>. After entering the blood, high S1P levels desensitize S1PR1, owing to phosphorylation by guanine nucleotide-binding protein coupled receptor kinase-2 (GRK2)<sup>68</sup>. Coordinated action of selectins, integrins and especially the retention signals of the CCL21 chemokine receptor CCR7 are required for lymphocytes to enter the lymph nodes through HEVs<sup>42</sup>. T lymphocytes sample the environment randomly — cortical sinus probing is followed by S1PR1-dependent entry<sup>69</sup> and then flow into the efferent lymph containing high S1P<sup>15</sup> (Fig. 4b). During inflammation, lymphocyte retention is increased by the loss of S1PR1 expression, its internalization and degradation. After several rounds of division, newly generated effector cells upregulate S1PR1, lose the CCR7 retention signal and then exit into circulation and travel to locations of injury and inflammation<sup>42</sup>. The immune system

co-opts the S1P–S1PR1 axis to determine the establishment of resident memory T cells, a distinct lineage that are embedded in non-lymphoid tissues and that mediate protective immunity. Regulation of the transcription factor KLF2 and its target gene *S1pr1* in mice by cytokines determines whether CD8<sup>+</sup> T cells re-enter circulation or become tissue resident<sup>70</sup>.

In addition to regulating the egress of B cells, the S1P–S1PR1 axis plays a part in the positioning of marginal-zone B cells in the spleen<sup>71</sup>. These B cells are the first line of defence against systemic blood-borne antigens. Levels of S1P are high in the splenic marginal zone, causing desensitization of S1PR1 on resident B cells that allows them to migrate into the follicles in response to the follicular-dendritic-cell-produced chemokine CXCL13 (ref. 42). B cells in the marginal zone shuttle constantly between this zone and follicles, depending on the S1P gradient. This allows for the rapid delivery of antigens from the blood to the follicles<sup>71</sup>. However, in germinal centres in which mature B cells proliferate, differentiate and switch antibody class production, S1PR2 activation and its downstream effectors G $\alpha$ 12, G $\alpha$ 13 and p115RhoGEF counter-act signalling by Akt, decreasing survival and responses to follicular chemoattractants, thus helping to focus the cells to the follicle centre<sup>43</sup>. It remains to be seen whether this gradient of S1P exists between the germinal centre and the follicle boundary. The remaining major challenges are to understand how this gradient is produced and how to measure S1P pools in local environments. For S1P to be sensed during egress, its levels in tissue interstitial fluid would be expected to be very low to ensure that B-cell S1PR1 could respond to higher concentrations at the luminal surfaces of egress sites. However, although interstitial levels are generally assumed to be due to high S1P degrading activity in tissues, so far it has not been possible to directly measure interstitial fluid S1P levels. It is anticipated that this will be accomplished soon with the improved sensitivity of newer mass spectrometers.

### S1P could be pro- or anti-inflammatory

The inflammatory role of S1P formed in immune cells by SphK1 or SphK2 is still unclear. This conundrum began when inflammatory responses were examined in *Sphk1* and *Sphk2* knockout mice. Whereas some studies have reported that colonic inflammation<sup>26</sup> and synovial inflammation in TNF-induced arthritis<sup>72</sup> are reduced in *Sphk1* knockout mice, others have reported that these mice have normal acute and chronic inflammatory responses<sup>73</sup>. The retraction of a paper that suggested SphK1 had a role in regulating pro-inflammatory responses in a mouse sepsis model was concerning<sup>74</sup>, and there is no evidence that inhibiting SphK1 or deleting its gene will reduce sepsis. Although *Sphk2* knockout mice have normal neutrophil functions, in a model of bacterial lung infection these mice have accelerated disease progression<sup>75</sup>. Similarly, neuroinflammation<sup>76</sup> and lung inflammatory injury<sup>77</sup> induced by lipopolysaccharide were increased by deletion of *Sphk1*.

It is possible that S1P formed by SphK1 or SphK2 has distinct functions and responses to stimuli in specific types of cells and tissues. This could also depend on the expression of S1P receptors and their specific downstream targets. Also, whereas deletion of *Sphk1* decreases S1P levels in blood, deletion of *Sphk2* increases it<sup>26,78</sup>, probably due to upregulation of *Sphk1* expression and an increase of S1P in some tissues but not others<sup>78</sup>. Changes in levels of S1P in the circulation and in tissues could affect immune-cell differentiation and functional phenotypes<sup>64</sup>. Production of S1P is spatially and temporally regulated and its physiological and inflammatory outcomes are context-, tissue- and source-dependent. Future studies with conditional knockout mice could clarify the inflammatory roles of S1P and the kinases that produce it in specific tissues.

### Sphingolipids may mediate inflammatory signalling by TNF

TNF, a key pro-inflammatory cytokine and therapeutic target in inflammatory diseases, stimulates sphingolipid metabolic enzymes including sphingomyelinases, ceramidase and SphK1 (refs 2, 79). Several sphingolipid metabolites have been suggested to be involved in TNF signalling and chemokine production based on studies of reduction or



overexpression of these enzymes. For example, cells that are deficient in acid sphingomyelinase or acid ceramidase have decreased production of CCL5 induced by TNF, possibly due to regulation of sphingosine levels<sup>23</sup>, whereas knockdown of the gene encoding neutral sphingomyelinase regulates adhesion molecules independently of NF- $\kappa$ B activation<sup>79</sup>. The effects of deletion or knockdown of *SphK1* have yielded conflicting results. Initially, overexpression of a dominant-negative SphK1 mutant<sup>80</sup> or SphK1 downregulation<sup>81,82</sup> reduced NF- $\kappa$ B activation, production of cytokines and chemokines, and induction of COX2 and PGE2 production in response to TNF. More recently, S1P formed by TNF-mediated activation of SphK1 was suggested to be crucial for TRAF2-mediated K63 polyubiquitylation of RIP1, a key step in NF- $\kappa$ B activation<sup>46</sup>. However, studies have now found that downregulation of SphK1 had no effect on TNF-induced NF- $\kappa$ B activation but enhanced CCL5 expression through the p38 MAPK pathway, whereas downregulation of SphK2 reduced CCL5 expression without affecting NF- $\kappa$ B<sup>83</sup>. Moreover, TNF-mediated NF- $\kappa$ B activation and cytokine expression in murine macrophages that lack both SphK1 and SphK2 were not altered but these cells had increased sphingosine and ceramide levels and autophagy<sup>24</sup>. Clearly, resolving these discrepancies requires further study; however, there are several intriguing possibilities. The interconversion of sphingolipid metabolites is rapid, and fluxes of sphingolipids depend on the cell type; thus, distinct bioactive sphingolipids could be increased by TNF and their targets could be cell-type dependent. The outcome of inside-out signalling by S1P induced by TNF, as observed in certain cell types<sup>3</sup>, would depend on the expression of S1P receptors and their downstream effectors. It is also possible that deletion of sphingosine kinases induces compensatory mechanisms. Moreover, TNF-induced inflammatory responses are not only regulated by NF- $\kappa$ B but are also controlled by crosstalk between NF- $\kappa$ B with other context-dependent pathways. Finally, TNF signalling leading to NF- $\kappa$ B and cytokine and chemokine production is now appreciated to be more complex than was originally suggested<sup>84</sup>. The recent discovery of RIP1-dependent and -independent activation of the early and late phases of I $\kappa$ B kinase (IKK), which are differentially regulated by low and high doses of TNF that lead to the expression of distinct sets of NF- $\kappa$ B target genes<sup>84</sup>, suggests a potential explanation for the divergent findings on the role of SphK1 and S1P. Exploring these possibilities might reconcile contradictory observations and shed new light on the inflammatory roles of S1P.

### Ceramides and inflammation

During host–pathogen defence, pattern recognition receptors, including Toll-like receptors (TLRs) and Nod-like receptors (NLRs), expressed by sentinel cells, macrophages and dendritic cells, bind microbial products to initiate innate immune responses and prime adaptive immunity. Integration of data from sphingolipidomics and transcriptomics of macrophages treated with lipopolysaccharide, the ligand for TLR4, revealed that several enzymes in sphingolipid metabolism and their products, particularly ceramide, were coordinately elevated<sup>85,86</sup>. Several intriguing ideas have been proposed to explain how ceramide links TLR4 and other members of this family to inflammation (Fig. 4). In macrophages, increased *de novo* ceramide biosynthesis is required for autophagosome formation<sup>86</sup>, which is thought to have key roles in innate immunity<sup>87</sup>. A question still to be answered is whether changes in autophagy owing to ceramide contributes to immunity and/or inflammation. Disruption of membrane microdomains in *Cers2* knockout mice, which are unable to make very-long-chain ceramides, prevented TNF-R1 internalization and downstream signalling<sup>88</sup>.

Another area that has received much attention is the connection between low-grade chronic inflammation and the pathogenesis of obesity and metabolic dysfunction<sup>89</sup>. A recently uncovered mechanistic link between increased inflammation and obesity is the activation of TLR4 by saturated fatty acids, leading to transcriptional activation of ceramide biosynthetic genes, including those that encode SPT and specific CerS isoforms, in an IKK- $\beta$ -dependent manner<sup>29</sup>. Interestingly, although not required for TLR4-dependent inflammatory cytokine induction, elevated ceramide production was required for TLR4-dependent insulin

resistance, owing to inhibition of Akt<sup>29</sup>, the activation of which is necessary for stimulated glucose transport<sup>90</sup> (Fig. 4). Likewise, *Cers2* knockout mice are glucose intolerant, owing to the inability of the insulin receptor to translocate to the disrupted membrane platforms, preventing Akt activation<sup>91</sup>. Together, these studies implicate ceramide as a mediator linking lipid-induced inflammatory pathways and insulin resistance.

The NLR family member NLRP3 inflammasome is activated by diverse non-microbial danger-associated molecular patterns (DAMPs) derived from damaged cells, and induces inflammation by increasing IL-1 $\beta$  and IL-18 secretion. In obese mice, lipotoxicity-associated ceramide was shown to increase activated caspase-1 in a Nlrp3 inflammasome-dependent manner in macrophages and adipose tissue<sup>92</sup>. This is the first suggestion that the Nlrp3 inflammasome senses increases in intracellular ceramide (Fig. 4)<sup>92,93</sup>. Moreover, age-related increased thymic ceramides also leads to Nlrp3 inflammasome-dependent caspase-1 activation<sup>94</sup>, suggesting that this form of ceramide elevation is responsible for the collapse of the thymic stromal cell microenvironment and subsequent decreased production of naive T cells and lower immune surveillance in the elderly<sup>94</sup>.

### Ceramides and adipokines

An intriguing potential link has surfaced between sphingolipid metabolism and the pleiotropic actions of the adipocyte hormone adiponectin. By binding to its receptors AdipoR1 or AdipoR2 adiponectin decreases inflammation and increases insulin sensitivity<sup>95</sup>. AdipoR1- and AdipoR2-stimulated ceramidase degraded ceramide to sphingosine, which was converted into the pro-survival S1P; evidence indicates that ceramidase activation initiates many adiponectin actions (Fig. 3)<sup>28</sup>. It is unclear whether AdipoR1 and AdipoR2 have intrinsic ceramidase activity, although a recent report suggests that AdipoR1, but not AdipoR2, induces caveolin-dependent recruitment of neutral ceramidase<sup>96</sup>.

Subsequent studies demonstrated that the adiponectin–ceramide axis is also involved in fibroblast growth factor-21 (FGF21)-mediated regulation of carbohydrate and lipid metabolism, and energy utilization. The beneficial effects of FGF21 were abolished in adiponectin-null mice, which were resistant to FGF21-induced decreases in ceramide and increases in energy expenditure<sup>97</sup>. Adiponectin also mediated the FGF21-induced decrease in production of liver pro-inflammatory cytokines in obese, diabetic mice<sup>98</sup>. Experiments are needed to determine which adiponectin-mediated effects depend mainly on decreasing ceramide signalling and which depend on increasing S1P signalling. Interestingly, thiazolidinediones — agonists for the nuclear receptor PPAR- $\gamma$ , widely used for management of type 2 diabetes — also induce FGF21 and adiponectin, and decrease ceramide in plasma and the liver<sup>97</sup>. These FGF21 and thiazolidinedione effects bring ceramides into the spotlight for research on new therapeutics for obesity-induced inflammation and metabolic disorders.

### New insights on the role of C1P in inflammation

C1P is a relatively new addition to the group of bioactive sphingolipid metabolites involved in inflammation. C1P activates cPLA2 $\alpha$ <sup>36</sup>, which releases arachidonic acid in the rate-determining step in eicosanoid production. As eicosanoids drive the pathogenesis of many inflammatory disorders, understanding the role of C1P in this process is crucial. An intriguing example of the latter was the characterization of a unique C1P-specific transfer protein called CPTP<sup>38</sup>. Depletion of CPTP, which is present in the cytosol and associated with the TGN and plasma membrane, increases C1P in the Golgi and nucleus, but decreases it in the plasma membrane. This suggested that CPTP controls levels of C1P produced by CERK in the Golgi by transferring it to the plasma membrane, thereby suppressing cPLA2 $\alpha$  activity and reducing arachidonic acid release and eicosanoid production<sup>38</sup>. The C1P that CPTP transfers to the plasma membrane is probably rapidly degraded by lipid phosphatases to ceramide. Another intriguing observation from this study was that changes in C1P levels directly correlated with changes in sphingosine and S1P, raising the possibility

that C1P might be a precursor of S1P that is mediated by, yet to be characterized, deacylase activity<sup>38</sup>. Global CERK inactivation has only modest effects on inflammatory responses, in contrast to *cPLA2* knockout, and further studies are necessary to address this discrepancy. Nevertheless, acute downregulation of CERK using short interfering RNA did reduce eicosanoid production in a mouse model of airway hyper-responsiveness<sup>99</sup>.

The concept of immunomodulation by C1P has become increasingly intricate as other studies have shown that C1P also negatively regulates inflammatory cytokine production. C1P produced by CERK regulates TNF processing and secretion by direct inhibition of TACE<sup>39</sup>. In addition, exogenous C1P was found to suppress lipopolysaccharide-mediated production of TNF, IL-6, IL-8 and IL-1 $\beta$  in human peripheral blood mononuclear cells<sup>100</sup>. Although exogenous C1P can stimulate macrophage migration by a pertussis-toxin-sensitive GPCR<sup>101</sup>, a cell surface C1P receptor has never been identified. It is also unclear what biological effects, if any, are regulated by circulating C1P.

### Therapeutic interventions in inflammatory disorders

Much of the clinical interest in sphingolipid metabolites and their signalling was spurred by the discovery of the mechanism of action of the immunosuppressant drug FTY720 (ref. 102). *In vivo*, SphK2 is mainly responsible for the phosphorylation of FTY720 to FTY720-P, an S1P mimetic and agonist for all S1P receptors except S1PR2. Binding of FTY720-P to S1PR1 leads to its downregulation, thereby effectively acting as a functional S1PR1 antagonist. Down-modulation of S1PR1 induces the retention of lymphocytes in lymphoid organs, termed lymphopenia, because the receptor is required for their exit into the blood up the S1P gradient; this prevents adaptive immune surveillance of tissues, and thus can ameliorate autoimmune diseases.

### Autoimmune diseases

Thousands of patients have been treated with FTY720 for the relapsing-remitting form of multiple sclerosis. In experimental autoimmune encephalomyelitis (EAE), an animal model of multiple sclerosis, FTY720 prevented pathogenic T cells committed to T helper (T<sub>H</sub>) 1 and T<sub>H</sub>17 lineages from crossing the blood-brain barrier into the central nervous system<sup>4</sup>. Down-modulation of S1PR1 in astrocytes by FTY720 reduced astrogliosis<sup>103</sup>, a hallmark of multiple sclerosis, and promoted remyelination and recovery of nerve conduction. These are likely explanations for the significant decrease in inflammatory markers and brain atrophy after FTY720 treatment of patients with multiple sclerosis<sup>4</sup>.

FTY720 and other next-generation S1PR1-modulating drugs are now in various stages of clinical trials for the treatment of a variety of immune diseases. One of the side effects of these modulating drugs is the transient reduction of heart rate by the activation of GIRK potassium channels downstream of cardiomyocyte S1PR1 (ref. 104), but with careful monitoring of patients after the first dose this should not prevent their use. FTY720 also induces pulmonary vascular leakage in mice by inducing degradation of S1PR1 (ref. 105). Receptor pharmacology predicts that competitive S1PR1 antagonists should not have this adverse effect; however, they could reduce barrier function, owing to S1PR1 blockade. So far, S1PR1 antagonists that are effective in various disease models, including EAE, have not been tested in the clinic. Even with FTY720's efficacy, the fact that it is well tolerated in patients and the convenience of oral administration, improvements could still be made. A deeper knowledge of S1P and its receptors, and an understanding of the beneficial and adverse side effects of targeting S1P receptors could lead to the development of even better treatments.

### Rheumatoid arthritis

Rheumatoid arthritis is an autoimmune disease characterized by inflammation in movable joints. High levels of S1P have been found in the synovial fluid of patients with rheumatoid arthritis<sup>30</sup>; however,

conflicting data were reported in different mouse rheumatoid arthritis models regarding the involvement of SphK1. Although *Sphk1* knockout mice showed normal responses in models of collagen-induced arthritis<sup>73</sup>, synovial inflammation in TNF-induced arthritis was reduced<sup>72</sup>. Hence, this area of research needs to be re-examined.

Lymphopenia-inducing drugs such as FTY720 have potential as rheumatoid arthritis therapeutics, FTY720 was protective in a mouse model of rheumatoid arthritis<sup>106</sup>; however, there are no current clinical trials. But, phase II clinical trials are ongoing to target S1P signalling by a completely different mode based on the finding that inhibition of SPL with the food dye THI (2-acetyl-4-tetrahydroxybutylimidazole) induces lymphopenia by elevating the levels of S1P in lymphoid tissues, disrupting the S1P gradient lymphocytes use for egress<sup>107</sup>. Subsequent studies of genetically deficient SPL mice revealed that they also had a disrupted S1P gradient and lymphopenia. On the basis of THI structure, a second compound, LX2931, was developed that was effective in a mouse model of rheumatoid arthritis and is now in phase II trials<sup>108</sup>. It remains to be seen whether direct inhibitors of SPL are clinically useful.

### Inflammatory bowel disease

Inflammatory bowel disease (IBD), a result of chronic inflammation of the gastrointestinal tract, is due to abnormal host immune responses to the intestinal microbiome<sup>109</sup>. Millions of people worldwide have ulcerative colitis and Crohn's disease, the major types of IBD. Studies using common acute and chronic epithelial injury colitis models have underscored a role for the SphK1–S1P–S1PR1 axis in IBD. Knockout of the gene encoding SphK1 or specific SphK1 inhibitors greatly reduce colitis severity, cytokine production and systemic inflammation<sup>26,78</sup>. In agreement, S1P is elevated during colitis and SphK1 expression is increased in patients with ulcerative colitis<sup>26</sup>. Several studies have also shown that FTY720 and other pharmacological agents that target S1PR1 are very effective in rodent models of colitis<sup>78,110,111</sup>, and several second-generation S1PR1 modulators are currently in clinical trials for patients with ulcerative colitis.

Persistent inflammation of the colon increases the risk of developing colorectal cancer<sup>112</sup>. There is increasing evidence from mouse models that SphK1 and S1P may be a link between chronic intestinal inflammation and the development of colitis-associated cancer (CAC), and that both are important for colon cancer development and progression<sup>78,113,114</sup>. An intracellular role for SphK1 in intestinal adenoma cell proliferation and polyp size was shown by deletion of SphK1 in *Apc*<sup>Min/+</sup> mice, a model of intestinal neoplasia<sup>113</sup>. Upregulation of SphK1 during colitis and CAC in mice leads to NF- $\kappa$ B-dependent production of IL-6, robust activation of the transcription factor Stat3, and subsequent upregulation of its target gene, *S1pr1*. FTY720 impeded the SphK1–S1P–S1PR1 axis, suppressing the NF- $\kappa$ B–IL-6–Stat3 vicious feed-forward amplification loop and CAC<sup>78</sup>. Thus, it seems that FTY720 and second-generation S1PR1 modulators should be considered for the treatment of inflammation-driven cancers.

### Asthma

Sphingolipid metabolites contribute to the pathogenesis of asthma, an increasingly prevalent chronic airway inflammation characterized by intermittent airflow obstruction and increased mucus production. S1P levels are elevated in the lungs of patients with allergic asthma. In mice, S1P is involved at multiple stages of the asthmatic response: it promotes the contraction of airway smooth muscle cells, the induction of airway hyper-reactivity by S1PR1 and/or S1PR3 (refs 115, 116), and regulates activation and functions of mast cells, eosinophils and dendritic cells<sup>27,117</sup>. Inhalation of an SphK1 inhibitor, or FTY720, mitigated asthma in rodent models<sup>27,117</sup>. *De novo* ceramide synthesis has also been implicated in human asthma based on the association of a 17q21 single nucleotide polymorphism (SNP) with increased expression of ORMDL3 (ref. 7), a negative regulator of



SPT, the first enzyme in sphingolipid synthesis<sup>6,118</sup>. Patients with this SNP are at increased risk of developing childhood asthma as well as virus-induced respiratory wheezing<sup>119</sup>. However, how ORMDL3 is involved in asthma remains a matter of debate. It has been shown that allergens upregulate ORMDL3 expression, and overexpression of ORMDL3 increased expression of asthma-associated chemokines and metalloproteases and the unfolded protein response (UPR), suggesting that ORMDL3 and asthma may be linked through the UPR pathway<sup>120</sup>. ORMDL3 also promotes eosinophil trafficking, recruitment and activation<sup>121</sup>. By contrast, although ceramide is elevated in murine models of asthma<sup>122</sup>, decreased *de novo* sphingolipid biosynthesis enhances bronchial reactivity by affecting intracellular magnesium homeostasis in the absence of inflammation<sup>123</sup>. Although these studies functionally link *de novo* sphingolipid biosynthesis to asthma, many questions remain unanswered, presenting challenging opportunities for research.

### Future perspectives

The study of bioactive sphingolipids in inflammation is only now coming to the foreground. As always in science, answering one question leads to many more, and this is especially true of these enigmatic sphingolipids. Clearly, more work is required to understand how inflammation regulates bioactive sphingolipid levels, the activities of the enzymes that control their levels and the proteins they signal through, and how this shapes the immune response. Although around a dozen extracellular and intracellular targets of signalling sphingolipids have been identified, it is unlikely that we have found them all, based on their pleiotropic effects. This is especially important for understanding the functions of ceramide in inflammation as only a few direct targets have been confirmed and how it regulates all inflammasomes remains a mystery. The complexity of sphingolipid metabolism and the 'sphingolipidome', and the fast interconversion of bioactive sphingolipids make this task even more daunting. This endeavour goes hand-in-hand with current systems-biology research, and it will be greatly aided by advanced sphingolipidomics — which can quantitatively measure changes in sphingolipids bound to their protein targets — combined with the development of conditional and tissue-specific sphingolipid-metabolic-enzyme knockout mice. Moreover, current sphingolipidomic technology for simultaneous quantification during inflammation of multiple sphingolipid species, for example, ceramides with different fatty acids, will be particularly helpful as these signalling molecules could have distinct and, sometimes, even opposing functions. Such a diverse array of bioactive sphingolipids could provide immune cells with a large toolkit to fine-tune specific inflammatory responses. Finally, improvement of matrix-assisted laser desorption ionization (MALDI) technology may, in the future, allow imaging of the S1P gradient in tissues such as the lymph nodes, the thymus and the spleen; this is crucial for our understanding of the role of the S1P–S1PR1 axis in lymphocyte trafficking.

Given the success of FTY720 for treatment of multiple sclerosis, it is hoped that next-generation S1PR1 modulators will find even wider therapeutic uses in other inflammatory disorders — as has been suggested by pre-clinical data. Progress in targeting other S1P receptors will increase our basic knowledge of their functions and how S1P can induce pro- and anti-inflammatory responses in different cell types, and will enable the rational design of more selective and potent inhibitors of immune responses. There are ongoing clinical trials with other interventions that target bioactive sphingolipids, such as Sphingomab, a monoclonal antibody that neutralizes S1P; ABC294640, which inhibits SphK2; and amitriptyline, which inhibits acid sphingomyelinase, and perhaps they will turn out to be useful for inflammatory disorders. Although such advances for other sphingolipid metabolites are still to be made, now that X-ray crystal structures of sphingolipid metabolic enzymes are available and carriers and targets of bioactive metabolites are gradually being

elucidated, it is expected that drugs will be developed to interfere with specific sphingolipid-mediated functions, which are important for the pathogenesis of human inflammatory disease. ■

Received 3 February; accepted 1 April 2014.

- Thudichum, J. L. W. *A Treatise on the Chemical Constitution of Brain* (Bailliere and Cox, 1884).
  - Hannun, Y. A. & Obeid, L. M. Principles of bioactive lipid signalling: lessons from sphingolipids. *Nature Rev. Mol. Cell Biol.* **9**, 139–150 (2008).
  - Spiegel, S. & Milstien, S. The outs and the ins of sphingosine-1-phosphate in immunity. *Nature Rev. Immunol.* **11**, 403–415 (2011).
  - Brinkmann, V. et al. Fingolimod (FTY720): discovery and development of an oral drug to treat multiple sclerosis. *Nature Rev. Drug Discov.* **9**, 883–897 (2010).
  - Garofalo, K. et al. Oral L-serine supplementation reduces production of neurotoxic deoxysphingolipids in mice and humans with hereditary sensory autonomic neuropathy type 1. *J. Clin. Invest.* **121**, 4735–4745 (2011).
  - Breslow, D. K. et al. Orm family proteins mediate sphingolipid homeostasis. *Nature* **463**, 1048–1053 (2010).
- This study was the first demonstration that Orm proteins are regulators of serine palmitoyltransferase, the first and rate-limiting step in sphingolipid synthesis, and provided key insights into how cells maintain sphingolipid homeostasis.**
- Moffatt, M. F. et al. A large-scale, consortium-based genome wide association study of asthma. *N. Engl. J. Med.* **363**, 1211–1221 (2010).
  - Hanada, K. Co-evolution of sphingomyelin and the ceramide transport protein CERT. *Biochim. Biophys. Acta* **1841**, 704–719 (2014).
  - D'Angelo, G., Rega, L. R. & De Matteis, M. A. Connecting vesicular transport with lipid synthesis: FAPP2. *Biochim. Biophys. Acta* **1821**, 1089–1095 (2012).
  - Dobrosotskaya, I. Y., Seegmiller, A. C., Brown, M. S., Goldstein, J. L. & Rawson, R. B. Regulation of SREBP processing and membrane lipid production by phospholipids in *Drosophila*. *Science* **296**, 879–883 (2002).
  - Nakahara, K. et al. The Sjogren-Larsson syndrome gene encodes a hexadecenal dehydrogenase of the sphingosine 1-phosphate degradation pathway. *Mol. Cell* **46**, 461–471 (2012).
  - Kharel, Y. et al. Sphingosine kinase type 1 inhibition reveals rapid turnover of circulating sphingosine 1-phosphate. *Biochem. J.* **440**, 345–353 (2011).
  - Pappu, R. et al. Promotion of lymphocyte egress into blood and lymph by distinct sources of sphingosine-1-phosphate. *Science* **316**, 295–298 (2007).
- The authors of this paper demonstrated that the S1P in blood is derived from haematopoietic cells, whereas lymph S1P was not, and clearly demonstrated that S1PR1 expression on lymphocytes was required for their egress.**
- Venkataraman, K. et al. Vascular endothelium as a contributor of plasma sphingosine 1-phosphate. *Circ. Res.* **102**, 669–676 (2008).
  - Pham, T. H. et al. Lymphatic endothelial cell sphingosine kinase activity is required for lymphocyte egress and lymphatic patterning. *J. Exp. Med.* **207**, 17–27 (2010).
  - Mendoza, A. et al. The transporter Spns2 is required for secretion of lymph but not plasma sphingosine-1-phosphate. *Cell Rep.* **2**, 1104–1110 (2012).
  - Fukuhara, S. et al. The sphingosine-1-phosphate transporter Spns2 expressed on endothelial cells regulates lymphocyte trafficking in mice. *J. Clin. Invest.* **122**, 1416–1426 (2012).
- Using global and endothelial-specific Spns2 knockout mice, these authors showed that Spns2 transports S1P out of endothelial cells and is required for B- and T-cell egress.**
- Quehenberger, O. et al. Lipidomics reveals a remarkable diversity of lipids in human plasma. *J. Lipid Res.* **51**, 3299–3305 (2010).
  - Christoffersen, C. et al. Endothelium-protective sphingosine-1-phosphate provided by HDL-associated apolipoprotein M. *Proc. Natl Acad. Sci. USA* **108**, 9613–9618 (2011).
  - Liu, M. et al. Hepatic ApoM overexpression stimulates formation of larger, ApoM/S1P-enriched plasma HDL. *J. Biol. Chem.* **289**, 2801–2814 (2014).
  - Kurano, M. et al. Liver involvement in sphingosine 1-phosphate dynamism revealed by adenoviral hepatic overexpression of apolipoprotein M. *Atherosclerosis* **229**, 102–109 (2013).
  - Suzuki, E., Handa, K., Toledo, M. S. & Hakomori, S. Sphingosine-dependent apoptosis: a unified concept based on multiple mechanisms operating in concert. *Proc. Natl Acad. Sci. USA* **101**, 14788–14793 (2004).
  - Jenkins, R. W. et al. Regulation of CC ligand 5/RANTES by acid sphingomyelinase and acid ceramidase. *J. Biol. Chem.* **286**, 13292–13303 (2011).
  - Xiong, Y. et al. Sphingosine kinases are not required for inflammatory responses in macrophages. *J. Biol. Chem.* **288**, 32563–32573 (2013).
  - Pitson, S. M. et al. Activation of sphingosine kinase 1 by ERK1/2-mediated phosphorylation. *EMBO J.* **22**, 5491–5500 (2003).
- This paper was the first demonstration that SphK1 is phosphorylated on Ser 225 by ERK, which is required for both activation and translocation of SphK1 to the plasma membrane.**
- Snider, A. J. et al. A role for sphingosine kinase 1 in dextran sulfate sodium-induced colitis. *FASEB J.* **23**, 143–152 (2009).
- This study was the first to show that S1P in colon and blood was increased in a mouse model of colitis.**
- Price, M. M. et al. A specific sphingosine kinase 1 inhibitor attenuates airway hyperresponsiveness and inflammation in a mast cell-dependent murine model of allergic asthma. *J. Allergy Clin. Immunol.* **131**, 501–511 (2013).

28. Holland, W. L. *et al.* Receptor-mediated activation of ceramidase activity initiates the pleiotropic actions of adiponectin. *Nature Med.* **17**, 55–63 (2011).  
**This was the first report showing that the anti-obesity and anti-inflammatory hormone adiponectin acts by stimulating ceramidase activity.**
29. Holland, W. L. *et al.* Lipid-induced insulin resistance mediated by the proinflammatory receptor TLR4 requires saturated fatty acid-induced ceramide biosynthesis in mice. *J. Clin. Invest.* **121**, 1858–1870 (2011).
30. Kitano, M. *et al.* Sphingosine 1-phosphate/sphingosine 1-phosphate receptor 1 signaling in rheumatoid synovium: regulation of synovial proliferation and inflammatory gene expression. *Arthritis Rheum.* **54**, 742–753 (2006); erratum **54**, 1704 (2006).
31. Majumdar, I. & Mastrandrea, L. D. Serum sphingolipids and inflammatory mediators in adolescents at risk for metabolic syndrome. *Endocrine* **41**, 442–449 (2012).
32. Canals, D., Roddy, P. & Hannun, Y. A. Protein phosphatase 1 $\alpha$  mediates ceramide-induced ERM protein dephosphorylation: a novel mechanism independent of phosphatidylinositol 4, 5-bisphosphate (PIP2) and myosin/ERM phosphatase. *J. Biol. Chem.* **287**, 10145–10155 (2012).
33. Saddoughi, S. A. *et al.* Sphingosine analogue drug FTY720 targets I2PP2A/SET and mediates lung tumour suppression via activation of PP2A-RIPK1-dependent necroptosis. *EMBO Mol. Med.* **5**, 105–121 (2013).
34. Fox, T. E. *et al.* Ceramide recruits and activates protein kinase C  $\zeta$  (PKC  $\zeta$ ) within structured membrane microdomains. *J. Biol. Chem.* **282**, 12450–12457 (2007).
35. Henry, B., Ziobro, R., Becker, K. A., Kolesnick, R. & Gulbins, E. Acid sphingomyelinase. *Handb. Exp. Pharmacol.* **215**, 77–88 (2013).
36. Pettus, B. J. *et al.* Ceramide-1-phosphate is a direct activator of cytosolic phospholipase A2. *J. Biol. Chem.* **279**, 11320–11326 (2004).  
**This study reported the first indication that C1P is a direct activator of cPLA2 $\alpha$ .**
37. Tauzin, L. *et al.* Effects of ceramide-1-phosphate on cultured cells: dependence on dodecane in the vehicle. *J. Lipid Res.* **48**, 66–76 (2007).
38. Simanshu, D. K. *et al.* Non-vesicular trafficking by a ceramide-1-phosphate transfer protein regulates eicosanoids. *Nature* **500**, 463–467 (2013).  
**This paper identified CPTP as a specific C1P transporter that transfers C1P from the Golgi and thereby negatively regulates C1P and eicosanoid production.**
39. Lamour, N. F. *et al.* Ceramide kinase regulates the production of tumor necrosis factor alpha (TNF $\alpha$ ) via inhibition of TNF $\alpha$ -converting enzyme. *J. Biol. Chem.* **286**, 42808–42817 (2011).
40. Lee, M. J. *et al.* Sphingosine-1-phosphate as a ligand for the G protein-coupled receptor EDG-1. *Science* **279**, 1552–1555 (1998).  
**This paper reported the first identification of a S1P receptor.**
41. Hla, T. & Dannenberg, A. J. Sphingolipid signaling in metabolic disorders. *Cell Metab.* **16**, 420–434 (2012).
42. Cyster, J. G. & Schwab, S. R. Sphingosine-1-phosphate and lymphocyte egress from lymphoid organs. *Annu. Rev. Immunol.* **30**, 69–94 (2012).
43. Green, J. A. *et al.* The sphingosine-1-phosphate receptor S1P $_2$  maintains the homeostasis of germinal center B cells and promotes niche confinement. *Nature Immunol.* **12**, 672–680 (2011).  
**The authors in this paper showed that S1P $_2$  expression on germinal centre B cells promotes apoptosis and inhibits migration, thus implicating S1P $_2$  as a negative regulator of B-cell activation and maturation.**
44. Takabe, K. & Spiegel, S. Export of sphingosine-1-phosphate and cancer progression. *J. Lipid Res.* <http://dx.doi.org/10.1194/jlr.R046656> (2014).
45. Hanson, M. A. *et al.* Crystal structure of a lipid G protein-coupled receptor. *Science* **335**, 851–855 (2012).  
**This was the first crystal structure of the lipid-binding GPCR S1PR1 that indicated that S1P enters the binding pocket by sliding in the plane of the membrane.**
46. Alvarez, S. E. *et al.* Sphingosine-1-phosphate is a missing cofactor for the E3 ubiquitin ligase TRAF2. *Nature* **465**, 1084–1088 (2010).  
**The authors of this paper demonstrated that S1P produced by SphK1 binds to and enhances the E3 ubiquitin ligase activity of TRAF2, an important component in NF- $\kappa$ B signalling induced by TNF.**
47. Harikumar, K. B. *et al.* K63-linked polyubiquitination of transcription factor IRF1 is essential for IL-1-induced production of chemokines CXCL10 and CCL5. *Nature Immunol.* **15**, 231–238 (2014).  
**This paper reported that in response to IL-1, cIAP2 and SphK1 form a complex with IRF1, leading to its activation by S1P-enhanced K63-linked polyubiquitylation, which in turn leads to the induction of IRF1-dependent chemokine genes that are important for sterile inflammation.**
48. Hait, N. C. *et al.* Regulation of histone acetylation in the nucleus by sphingosine-1-phosphate. *Science* **325**, 1254–1257 (2009).
49. Ihlefeld, K., Claas, R. F., Koch, A., Pfeilschifter, J. M. & Meyer Zu Heringdorf, D. Evidence for a link between histone deacetylation and Ca $^{2+}$  homeostasis in sphingosine-1-phosphate lyase-deficient fibroblasts. *Biochem. J.* **447**, 457–464 (2012).
50. Nguyen-Tran, D. H. *et al.* Molecular mechanism of sphingosine-1-phosphate action in Duchenne muscular dystrophy. *Dis. Model. Mech.* **7**, 41–54 (2014).
51. Petrache, I., Petrusca, D. N., Bowler, R. P. & Kamocki, K. Involvement of ceramide in cell death responses in the pulmonary circulation. *Proc. Am. Thorac. Soc.* **8**, 492–496 (2011).
52. Lee, M. J. *et al.* Vascular endothelial cell adherens junction assembly and morphogenesis induced by sphingosine-1-phosphate. *Cell* **99**, 301–312 (1999).  
**This paper reported that S1P regulates adherens junctions.**
53. Garcia, J. G. *et al.* Sphingosine-1-phosphate promotes endothelial cell barrier integrity by Edg-dependent cytoskeletal rearrangement. *J. Clin. Invest.* **108**, 689–701 (2001).
54. Natarajan, V. *et al.* Sphingosine-1-phosphate, FTY720, and sphingosine-1-phosphate receptors in the pathobiology of acute lung injury. *Am. J. Respir. Cell Mol. Biol.* **49**, 6–17 (2013).
55. Zhang, Q. J. *et al.* Ceramide mediates vascular dysfunction in diet-induced obesity by PP2A-mediated dephosphorylation of the eNOS-Akt complex. *Diabetes* **61**, 1848–1859 (2012).  
**The authors of this article provided evidence that obesity-induced endothelial dysfunction is due to elevation of ceramide, which de-represses PP2A, leading to dephosphorylation and inhibition of Akt and eNOS.**
56. Predescu, S. *et al.* Platelet activating factor-induced ceramide micro-domains drive endothelial NOS activation and contribute to barrier dysfunction. *PLoS ONE* **8**, e75846 (2013).
57. Camerer, E. *et al.* Sphingosine-1-phosphate in the plasma compartment regulates basal and inflammation-induced vascular leak in mice. *J. Clin. Invest.* **119**, 1871–1879 (2009).
58. Herzog, B. H. *et al.* Podoplanin maintains high endothelial venule integrity by interacting with platelet CLEC-2. *Nature* **502**, 105–109 (2013).
59. Kumaraswamy, S. B., Linder, A., Akesson, P. & Dahlback, B. Decreased plasma concentrations of apolipoprotein M in sepsis and systemic inflammatory response syndromes. *Crit. Care* **16**, R60 (2012).
60. Cui, H. *et al.* Sphingosine-1-phosphate receptor 2 protects against anaphylactic shock through suppression of endothelial nitric oxide synthase in mice. *J. Allergy Clin. Immunol.* **132**, 1205–1214 (2013).
61. Zhang, G. *et al.* Critical role of sphingosine-1-phosphate receptor 2 (S1PR2) in acute vascular inflammation. *Blood* **122**, 443–455 (2013).
62. Teijaro, J. R. *et al.* Endothelial cells are central orchestrators of cytokine amplification during influenza virus infection. *Cell* **146**, 980–991 (2011).
63. Teijaro, J. R., Walsh, K. B., Rice, S., Rosen, H. & Oldstone, M. B. Mapping the innate signaling cascade essential for cytokine storm during influenza virus infection. *Proc. Natl Acad. Sci. USA* **111**, 3799–3804 (2014).
64. Rivera, J., Proia, R. L. & Olivera, A. The alliance of sphingosine-1-phosphate and its receptors in immunity. *Nature Rev. Immunol.* **8**, 753–763 (2008).
65. Chi, H. Sphingosine-1-phosphate and immune regulation: trafficking and beyond. *Trends Pharmacol. Sci.* **32**, 16–24 (2011).
66. Takada, K. *et al.* Kruppel-like factor 2 is required for trafficking but not quiescence in postactivated T cells. *J. Immunol.* **186**, 775–783 (2011).
67. Faroudi, M. *et al.* Critical roles for Rac GTPases in T-cell migration to and within lymph nodes. *Blood* **116**, 5536–5547 (2010).
68. Arnon, T. I. *et al.* GRK2-dependent S1PR1 desensitization is required for lymphocytes to overcome their attraction to blood. *Science* **333**, 1898–1903 (2011).  
**The authors of this paper explain how lymphocytes migrate into lymphoid tissues against the S1P gradient by GRK2-dependent downregulation of S1PR1.**
69. Grigorova, I. L., Panteleev, M. & Cyster, J. G. Lymph node cortical sinus organization and relationship to lymphocyte egress dynamics and antigen exposure. *Proc. Natl Acad. Sci. USA* **107**, 20447–20452 (2010).
70. Skon, C. N. *et al.* Transcriptional downregulation of S1pr1 is required for the establishment of resident memory CD8 $^{+}$  T cells. *Nature Immunol.* **14**, 1285–1293 (2013).
71. Arnon, T. I., Horton, R. M., Grigorova, I. L. & Cyster, J. G. Visualization of splenic marginal zone B-cell shuttling and follicular B-cell egress. *Nature* **493**, 684–688 (2013).
72. Baker, D. A., Barth, J., Chang, R., Obeid, L. M. & Gilkeson, G. S. Genetic sphingosine kinase 1 deficiency significantly decreases synovial inflammation and joint erosions in murine TNF- $\alpha$ -induced arthritis. *J. Immunol.* **185**, 2570–2579 (2010).
73. Michaud, J., Kohno, M., Proia, R. L. & Hla, T. Normal acute and chronic inflammatory responses in sphingosine kinase 1 knockout mice. *FEBS Lett.* **580**, 4607–4612 (2006).
74. Puneet, P. *et al.* Retraction. *Science* **341**, 342 (2013).
75. Zemmann, B. *et al.* Normal neutrophil functions in sphingosine kinase type 1 and 2 knockout mice. *Immunol. Lett.* **109**, 56–63 (2007).
76. Grin'kina, N. M. *et al.* Sphingosine kinase 1 deficiency exacerbates LPS-induced neuroinflammation. *PLoS ONE* **7**, e36475 (2012).
77. Di, A. *et al.* A novel function of sphingosine kinase 1 suppression of JNK activity in preventing inflammation and injury. *J. Biol. Chem.* **285**, 15848–15857 (2010).
78. Liang, J. *et al.* Sphingosine-1-phosphate links persistent STAT3 activation, chronic intestinal inflammation, and development of colitis-associated cancer. *Cancer Cell* **23**, 107–120 (2013).
79. Clarke, C. J., Truong, T. G. & Hannun, Y. A. Role for neutral sphingomyelinase-2 in tumor necrosis factor alpha-stimulated expression of vascular cell adhesion molecule-1 (VCAM) and intercellular adhesion molecule-1 (ICAM) in lung epithelial cells: p38 MAPK is an upstream regulator of nSMase2. *J. Biol. Chem.* **282**, 1384–1396 (2007).  
**This paper is the first demonstration of the involvement of neutral sphingomyelinase in TNF actions.**
80. Xia, P. *et al.* Sphingosine kinase interacts with TRAF2 and dissects tumor necrosis factor- $\alpha$  signaling. *J. Biol. Chem.* **277**, 7996–8003 (2002).
81. Billich, A. *et al.* Basal and induced sphingosine kinase 1 activity in A549 carcinoma cells: function in cell survival and IL-1 $\beta$  and TNF- $\alpha$  induced production of inflammatory mediators. *Cell. Signal.* **17**, 1203–1217 (2005).



82. Pettus, B. J. *et al.* The sphingosine kinase 1/sphingosine-1-phosphate pathway mediates COX-2 induction and PGE2 production in response to TNF- $\alpha$ . *FASEB J.* **17**, 1411–1421 (2003).
83. Adada, M. M. *et al.* Sphingosine kinase 1 regulates tumor necrosis factor- $\alpha$ -induced RANTES induction through p38 mitogen-activated protein kinase but independently of nuclear factor  $\kappa$ B activation. *J. Biol. Chem.* **288**, 27667–27679 (2013).
84. Blackwell, K. *et al.* Two coordinated mechanisms underlie tumor necrosis factor  $\alpha$ -induced immediate and delayed I $\kappa$ B kinase activation. *Mol. Cell. Biol.* **33**, 1901–1915 (2013).
85. Dennis, E. A. *et al.* A mouse macrophage lipidome. *J. Biol. Chem.* **285**, 39976–39985 (2010).
86. Sims, K. *et al.* Kdo2-lipid A, a TLR4-specific agonist, induces *de novo* sphingolipid biosynthesis in RAW264.7 macrophages, which is essential for induction of autophagy. *J. Biol. Chem.* **285**, 38568–38579 (2010).
87. Levine, B., Mizushima, N. & Virgin, H. W. Autophagy in immunity and inflammation. *Nature* **469**, 323–335 (2011).
88. Ali, M. *et al.* Altering the sphingolipid acyl chain composition prevents LPS/GLN-mediated hepatic failure in mice by disrupting TNFR1 internalization. *Cell Death Dis.* **4**, e929 (2013).
89. Hotamisligil, G. S. Inflammation and metabolic disorders. *Nature* **444**, 860–867 (2006).
90. Chavez, J. A. & Summers, S. A. A ceramide-centric view of insulin resistance. *Cell Metab.* **15**, 585–594 (2012).
91. Park, J. W. *et al.* Ablation of very long acyl chain sphingolipids causes hepatic insulin resistance in mice due to altered detergent-resistant membranes. *Hepatology* **57**, 525–532 (2013).
92. Vandanmagsar, B. *et al.* The NLRP3 inflammasome instigates obesity-induced inflammation and insulin resistance. *Nature Med.* **17**, 179–188 (2011).  
**This report is the first demonstration that ceramide can activate an inflammasome, leading to caspase-1 cleavage.**
93. Youm, Y. H. *et al.* Canonical Nlrp3 inflammasome links systemic low-grade inflammation to functional decline in aging. *Cell Metab.* **18**, 519–532 (2013).
94. Youm, Y. H. *et al.* The Nlrp3 inflammasome promotes age-related thymic demise and immunosenescence. *Cell Rep.* **1**, 56–68 (2012).
95. Yamauchi, T. *et al.* Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase. *Nature Med.* **8**, 1288–1295 (2002).
96. Wang, Y. *et al.* Adiponectin inhibits tumor necrosis factor- $\alpha$ -induced vascular inflammatory response via caveolin-mediated ceramidase recruitment and activation. *Circ. Res.* **114**, 792–805 (2014).
97. Holland, W. L. *et al.* An FGF21-adiponectin-ceramide axis controls energy expenditure and insulin action in mice. *Cell Metab.* **17**, 790–797 (2013).  
**The authors of this paper showed that the anti-obesity hormone FGF21 acts through stimulation of adiponectin secretion that in turn decreases ceramide levels.**
98. Lin, Z. *et al.* Adiponectin mediates the metabolic effects of FGF21 on glucose homeostasis and insulin sensitivity in mice. *Cell Metab.* **17**, 779–789 (2013).
99. Miettla, J. A. *et al.* Characterization of eicosanoid synthesis in a genetic ablation model of ceramide kinase. *J. Lipid Res.* **54**, 1834–1847 (2013).
100. Hankins, J. L., Fox, T. E., Barth, B. M., Unrath, K. A. & Kester, M. Exogenous ceramide-1-phosphate reduces lipopolysaccharide (LPS)-mediated cytokine expression. *J. Biol. Chem.* **286**, 44357–44366 (2011).
101. Granado, M. H. *et al.* Ceramide 1-phosphate (C1P) promotes cell migration. Involvement of a specific C1P receptor. *Cell. Signal.* **21**, 405–412 (2009).
102. Mandala, S. *et al.* Alteration of lymphocyte trafficking by sphingosine-1-phosphate receptor agonists. *Science* **296**, 346–349 (2002).  
**This is the first demonstration that the phosphorylated form of FTY720 is an agonist for all S1P receptors, but S1PR2 and that FTY720-P induces lymphopenia.**
103. Choi, J. W. *et al.* FTY720 (fingolimod) efficacy in an animal model of multiple sclerosis requires astrocyte sphingosine 1-phosphate receptor 1 (S1P1) modulation. *Proc. Natl Acad. Sci. USA* **108**, 751–756 (2011).  
**This study identified non-immunological central nervous system mechanisms of FTY720 efficacy in EAE that depend on the drugs ability to modulate S1PR1 expression in astrocytes.**
104. Gergely, P. *et al.* The selective S1P receptor modulator BAF312 redirects lymphocyte distribution and has species-specific effects on heart rate: translation from preclinical to clinical studies. *Br. J. Pharmacol.* **167**, 1035–1047 (2012).
105. Oo, M. L. *et al.* Engagement of S1P<sub>1</sub>-degradative mechanisms leads to vascular leak in mice. *J. Clin. Invest.* **121**, 2290–2300 (2011).
106. Tsunemi, S. *et al.* Effects of the novel immunosuppressant FTY720 in a murine rheumatoid arthritis model. *Clin. Immunol.* **136**, 197–204 (2010).
107. Schwab, S. R. *et al.* Lymphocyte sequestration through S1P lyase inhibition and disruption of S1P gradients. *Science* **309**, 1735–1739 (2005).  
**This paper provides compelling evidence that lymphocyte egress depends on the S1P gradient.**
108. Bagdanoff, J. T. *et al.* Inhibition of sphingosine 1-phosphate lyase for the treatment of rheumatoid arthritis: discovery of (E)-1-(4-((1R,2S,3R)-1,2,3,4-tetrahydroxybutyl)-1H-imidazol-2-yl)ethanone oxime (LX2931) and (1R,2S,3R)-1-(2-(isoxazol-3-yl)-1H-imidazol-4-yl)butane-1,2,3,4-tetraol (LX2932). *J. Med. Chem.* **53**, 8650–8662 (2010).
109. Khor, B., Gardet, A. & Xavier, R. J. Genetics and pathogenesis of inflammatory bowel disease. *Nature* **474**, 307–317 (2011).
110. Daniel, C. *et al.* FTY720 ameliorates oxazolone colitis in mice by directly affecting T helper type 2 functions. *Mol. Immunol.* **44**, 3305–3316 (2007).
111. Sanada, Y. *et al.* Therapeutic effects of novel sphingosine-1-phosphate receptor agonist W-061 in murine DSS colitis. *PLoS ONE* **6**, e23933 (2011).
112. Ullman, T. A. & Itzkowitz, S. H. Intestinal inflammation and cancer. *Gastroenterology* **140**, 1807–1816 (2011).
113. Kohno, M. *et al.* Intracellular role for sphingosine kinase 1 in intestinal adenoma cell proliferation. *Mol. Cell. Biol.* **26**, 7211–7223 (2006).
114. Kawamori, T. *et al.* Role for sphingosine kinase 1 in colon carcinogenesis. *FASEB J.* **23**, 405–414 (2009).
115. Trifilieff, A. & Fozard, J. R. Sphingosine-1-phosphate-induced airway hyper-reactivity in rodents is mediated by the sphingosine-1-phosphate type 3 receptor. *J. Pharmacol. Exp. Ther.* **342**, 399–406 (2012).
116. Olivera, A. *et al.* Sphingosine-1-phosphate can promote mast cell hyper-reactivity through regulation of contactin-4 expression. *J. Leukoc. Biol.* **94**, 1013–1024 (2013).
117. Nishiuma, T. *et al.* Inhalation of sphingosine kinase inhibitor attenuates airway inflammation in asthmatic mouse model. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **294**, L1085–L1093 (2008).
118. Siow, D. L. & Wattenberg, B. W. Mammalian ORMDL proteins mediate the feedback response in ceramide biosynthesis. *J. Biol. Chem.* **287**, 40198–40204 (2012).
119. Çalıskan, M. *et al.* Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. *N. Engl. J. Med.* **368**, 1398–1407 (2013).
120. Miller, M. *et al.* ORMDL3 is an inducible lung epithelial gene regulating metalloproteases, chemokines, OAS, and ATF6. *Proc. Natl Acad. Sci. USA* **109**, 16648–16653 (2012).
121. Ha, S. G. *et al.* ORMDL3 promotes eosinophil trafficking and activation via regulation of integrins and CD48. *Nature Commun.* **4**, 2479 (2013).
122. Masini, E. *et al.* Ceramide: a key signaling molecule in a Guinea pig model of allergic asthmatic response and airway inflammation. *J. Pharmacol. Exp. Ther.* **324**, 548–557 (2008).
123. Worgall, T. S. *et al.* Impaired sphingolipid synthesis in the respiratory tract induces airway hyperreactivity. *Sci. Transl. Med.* **5**, 186ra67 (2013).  
**This work showed that inhibition of sphingolipid biosynthesis enhances airway hyper-responsiveness.**

**Acknowledgments** We thank S. Milstien for critically reading the manuscript. This work was supported by grants R37GM043880, R01AI500941 and R01CA61774 from the US National Institutes of Health. We apologize to those whose work was not cited owing to space constraints.

**Author Information** Reprints and permissions information are available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at [go.nature.com/znpg6g](http://go.nature.com/znpg6g). Correspondence should be addressed to S.S. ([sspiegel@vcu.edu](mailto:sspiegel@vcu.edu)).

# Sphingolipid lysosomal storage disorders

Frances M. Platt<sup>1</sup>

**Lysosomal storage diseases are inborn errors of metabolism, the hallmark of which is the accumulation, or storage, of macromolecules in the late endocytic system. They are monogenic disorders that occur at a collective frequency of 1 in 5,000 live births and are caused by inherited defects in genes that mainly encode lysosomal proteins, most commonly lysosomal enzymes. A subgroup of these diseases involves the lysosomal storage of glycosphingolipids. Through our understanding of the genetics, biochemistry and, more recently, cellular aspects of sphingolipid storage disorders, we have gained insights into fundamental aspects of cell biology that would otherwise have remained opaque. In addition, study of these disorders has led to significant progress in the development of therapies, several of which are now in routine clinical use. Emerging mechanistic links with more common diseases suggest we need to rethink our current concept of disease boundaries.**

Lipids, in common with other cellular macromolecules, have a shared fate. They undergo catabolism in late endosomes/lysosomes through the action of acid hydrolases. In this Review, I focus on sphingolipids, and in particular on glycosphingolipids (GSLs), and the diseases that arise through defects in their catabolism — the glycosphingolipidoses<sup>1</sup>. The glycosphingolipidoses are a prototypical family of lysosomal storage diseases (LSDs) that share some features with other lysosomal disorders<sup>2</sup>. Indeed, many of the therapeutic approaches for treating LSDs were originally developed for treating sphingolipid lysosomal diseases.

GSLs are a complex group of sphingolipids, in which the ceramide backbone is modified by the addition of a carbohydrate head group. The simplest GSLs have a single monohexoside (glucose or galactose) moiety linked to ceramide<sup>3</sup>. Over evolutionary time the carbohydrate head group has increased in complexity — more than 300 structures have been described, with those derived from the core structure glucosylceramide (GlcCer) being the most common. Galactosylceramide (GalCer) and its sulphated derivative have a much more restricted distribution and are generally confined to myelin and the kidney<sup>3</sup>. When the oligosaccharide head group of GlcCer-derived GSLs contains a sialic acid, the GSL is charged, and these GSLs are termed gangliosides and are the major glycoconjugates found in the nervous system<sup>4</sup> (Fig. 1).

The complexity of sphingolipids is not just the result of the glycan head group, but also the ceramide chain-length heterogeneity that arises owing to the differential chain-length specificity of ceramide synthases<sup>5</sup>.

## Inherited diseases of GSL biosynthesis and catabolism

Inborn errors of GSL metabolism occur in the human population at low frequency. They result in severe monogenic diseases, most typically inherited as autosomal recessive traits. For many years it was thought that no GSL biosynthetic diseases existed owing to presumed *in utero* lethality. Loss of GSL biosynthesis by targeted disruption of GlcCer synthase (that catalyses the first steps in GSL biosynthesis) results in lethality with a failure of embryos to develop beyond gastrulation in mice<sup>6</sup>. However, engineered mice that lack key enzymes involved in downstream steps of GSL biosynthesis, namely ganglioside biosynthesis, are viable<sup>7</sup>. This suggests that loss

of the entire GlcCer biosynthetic pathway is incompatible with normal development, but loss of certain complex ganglioside structures is tolerated, albeit with accompanying neurological dysfunction in some cases<sup>7</sup>.

## Biosynthetic diseases

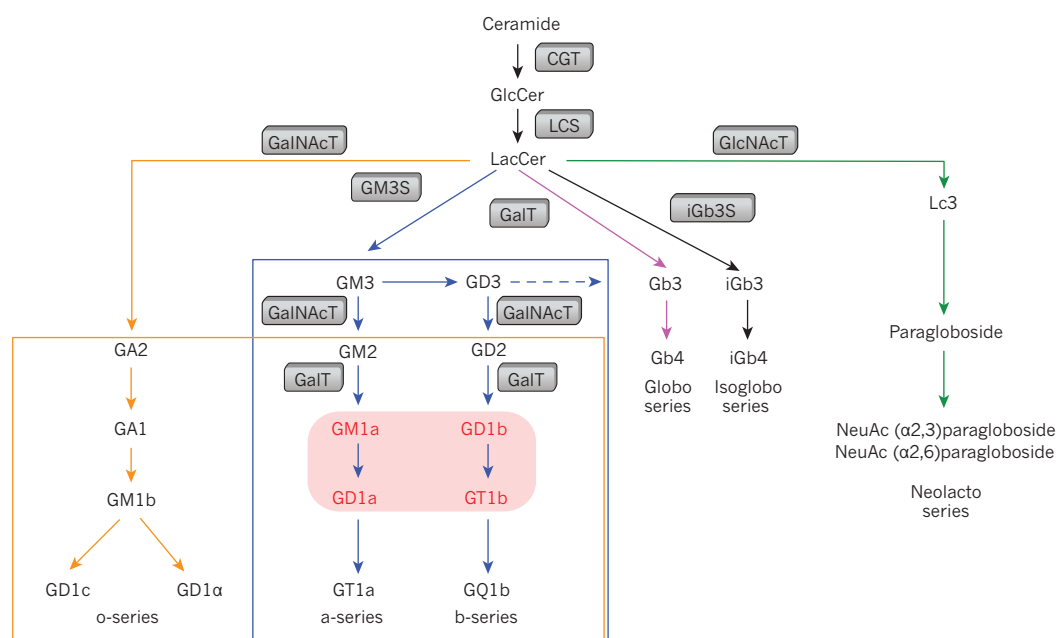
Two disorders of ganglioside biosynthesis have been genetically proven in the human population within the past decade, and result from GM3-synthase<sup>8</sup> and GM2-synthase deficiency<sup>9,10</sup> (Table 1 and Fig. 1). GM3-synthase deficiency, which was discovered in the Old Order Amish community in Ohio, presents as a severe epilepsy syndrome. By contrast, GM2-synthase deficiency results in spastic paraplegia (progressive lower limb weakness and spasticity). When GSL levels were measured in patient plasma, the predicted deficiencies in GM2 or GM3, and their downstream derivatives, were confirmed<sup>8,9</sup>. However, levels of the precursors of GM2 or GM3 were elevated owing to the block in ganglioside biosynthesis<sup>8,9</sup> (Fig. 1). Furthermore, GM2-synthase deficiency resulted in the biosynthesis of minute amounts of a novel GSL structure (sialylated Gb3) that has not been previously reported in biological systems<sup>9</sup>. It is currently unclear whether these diseases are caused by the loss of the specific ganglioside species, by elevation in precursor GSLs or by an imbalance in the charged/neutral GSL ratios<sup>11</sup>. These possibilities are not mutually exclusive. However, knowledge of the precise biochemical or pathophysiological basis for these diseases will be crucial for the development of therapeutic strategies for these severe, currently untreatable metabolic diseases. It is highly probable that these two disorders are not the only diseases that result from defects in complex ganglioside biosynthesis, and it is likely that others will be identified within the human population and attributed to defects in other enzymes involved in ganglioside biosynthesis (Fig. 1).

## Catabolic diseases

In contrast to the paucity of GSL biosynthetic diseases, there is a relatively large group of diseases that result from defects in GSL catabolism<sup>12</sup> (Table 1). These diseases constitute an important subgroup within a larger family of more than 70 LSDs. There has been debate in the field as to whether the term 'storage' is really appropriate for these

<sup>1</sup>Department of Pharmacology, University of Oxford, Oxford OX1 3QT, UK.





**Figure 1 | Glycosphingolipid biosynthetic pathway.** The gangliosides lost in GM2 synthase deficiency are indicated by the orange box and the gangliosides lost in GM3 deficiency are indicated by the blue box. The overlap of the orange and blue boxes indicates common gangliosides lost in both diseases. The four major gangliosides of the mammalian brain

are shown in red. The key enzymes ceramide glucosyl transferase (CGT), *N*-acetylgalactosamine transferase (GalNAcT), galactosyl transferase (GalT), *N*-acetylglucosamine transferase (GlcNAcT), GM3 synthase (GM3S), iGb3 synthase (iGb3S) and lactosyl ceramide synthase (LCS) are shown in grey.

diseases. It perhaps over emphasizes one aspect of these complex, multifaceted disorders, and could be misleading when considering pathogenesis. However, at present, no consensus on any nomenclature change has been debated or agreed. In this Review, I therefore refer to LSDs when discussing monogenic catabolic diseases and use the term 'lysosomal diseases' to encompass any disease with lysosomal dysfunction, irrespective of the cause.

Many LSDs have a progressive neurodegenerative clinical course and often present in infancy or childhood<sup>13</sup>; however, symptom onset can occur at any stage of life and late-onset variants have also been described<sup>14–16</sup>. A major factor that has promoted our understanding

of pathogenesis, and is a platform for testing therapies, is the availability of a large number of authentic animal models for these disorders<sup>17,18</sup>. These models span mutant and engineered mice through to spontaneous mutants described in species such as domestic cats, dogs, sheep, cattle<sup>17</sup>, and even a flamingo with Tay–Sachs disease<sup>19</sup>. The availability of large and small animal models has catalysed considerable progress in our understanding of these disorders, as well as in the development of therapies for specific conditions<sup>20–23</sup>. This is in contrast to more common sporadic neurodegenerative diseases, such as Alzheimer's disease. In Alzheimer's disease the lack of authentic animal models is a significant factor that contributes to the dearth

**Table 1 | Inborn errors of GLS metabolism that involve GlcCer-derived GSLs**

Metabolic defect	Disease	Biochemical defect	Gene involved	Protein defect	Major symptoms	CNS pathology	References
Biosynthetic	GM3 synthase deficiency	Loss of GM3 and downstream gangliosides	<i>ST3GAL5</i>	GM3S	Infantile onset epilepsy with developmental arrest and blindness	Yes	8
Biosynthetic	GM2 synthase deficiency	Loss of GM2 and downstream gangliosides	<i>B4GALNT1</i>	GM2S	Spastic paraplegia	Yes	9, 10
Catabolic (LSD)	Gaucher types 1, 2 and 3	GlcCer storage	<i>GBA</i>	Glucocerebrosidase	Hepatosplenomegaly, haematological defects, inflammation, bone disease and CNS disease in types 2 and 3	In types 2 and 3	43, 108, 109
Catabolic (LSD)	Fabry	Gb3 storage	<i>GLA</i>	$\alpha$ -galactosidase	Renal, cardiovascular and peripheral pain	Yes (some mild cases of cerebrovasculopathy)	110, 111
Catabolic (LSD)	Tay–Sachs	GM2 ganglioside storage	<i>HEXA</i>	B-hexosaminidase $\alpha$ subunit	Progressive neurodegeneration	Yes	112, 113, 114
Catabolic (LSD)	Sandhoff	GM2 ganglioside storage	<i>HEXB</i>	B-hexosaminidase $\beta$ subunit	Progressive neurodegeneration	Yes	115, 116
Catabolic (LSD)	GM1 gangliosidosis	GM1 ganglioside storage	<i>GLB1</i>	$\beta$ -galactosidase	Progressive neurodegeneration	Yes	117, 118, 119
Defect in trafficking or fusion	Niemann–Pick type C	Storage of all GSLs, cholesterol, sphingomyelin and sphingosine	<i>NPC1</i> <i>NPC2</i>	NPC1 NPC2	Progressive neurodegeneration	Yes	120, 121

CNS, central nervous system; GlcCer, glucosylceramide; GM2S, GM2 synthase; GM3S, GM3 synthase; GSL, glycosphingolipid; LSD, lysosomal storage disease.

of treatment options available for this common disease of ageing<sup>24</sup>. In the remainder of this article I focus on recent advances in our understanding and treatment of GSL LSDs.

## GSL LSDs

Lysosomal diseases that involve sphingolipid storage occur through two distinct mechanisms<sup>25</sup>. Primary sphingolipidoses are caused by inherited defects in genes encoding the lysosomal enzymes (or their non-enzymatic cofactors) required for sphingolipid catabolism in late endosomes/lysosomes (Table 1 and Fig. 2). However, there are also LSDs involving sphingolipid storage in which the primary defect is not in a catabolic enzyme but in, for example, a lysosomal membrane protein<sup>25</sup>. LSDs have brought to light a number of proteins with unknown function that reside in the limiting membrane of the lysosome<sup>26</sup>. Frequently these diseases are associated with defects in trafficking and fusion in the endocytic system, leading to secondary storage of sphingolipids<sup>25</sup>. An extreme example of this is Niemann–Pick type C (NPC), a disease thought to be mainly a disorder of cholesterol transport from the lysosome<sup>27</sup>. However, sphingolipids are also implicated in the aetiology of this disease because impaired sphingosine efflux from the lysosome results in incomplete filling of the acidic store with calcium; this, in turn, results in reduced calcium release, which blocks late-endosome–lysosome fusion<sup>28</sup>. This causes secondary storage of GSLs in late endosomes/lysosomes as well as other lipids, including low-density lipoprotein (LDL)-derived cholesterol<sup>29</sup>. Furthermore, the only approved therapy (miglustat) for NPC disease is an inhibitor of GSL biosynthesis<sup>30</sup>, suggesting that sphingolipids have a role in this disease.

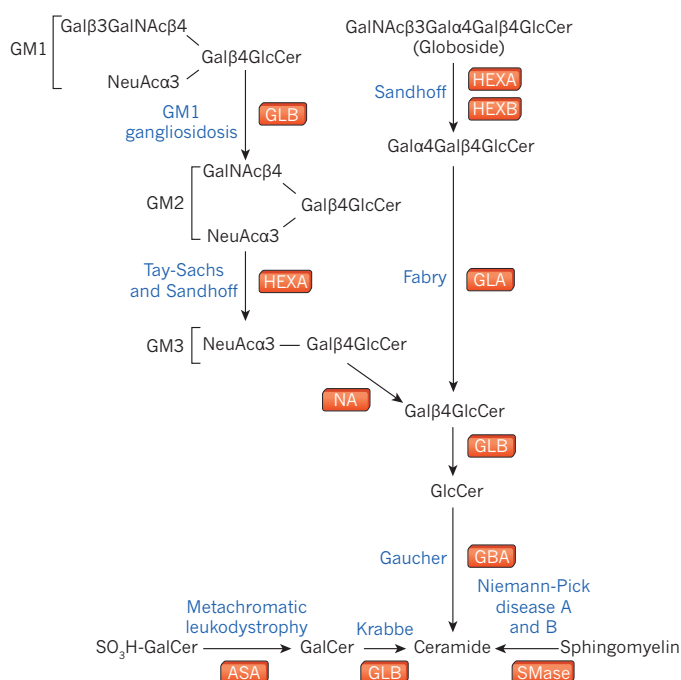
## Current therapies for GSL LSDs

The only potential cure for any monogenic disease is gene therapy<sup>31</sup>. LSDs are seen as attractive candidates for gene therapy because lysosomal enzymes are secreted and recaptured by neighbouring cells so that not every cell needs transducing. These diseases are therefore viewed as tractable for the application of this therapeutic approach. So far, this form of therapy has not yet reached routine clinical practice. However,

proof-of-concept studies have shown efficacy in authentic small- and large-animal disease models<sup>32–36</sup>, leading to clinical trials (Table 2). If these trials are successful, the challenge will be to intervene early, before irreversible pathology has occurred<sup>37,38</sup>. Another key factor is that the gene delivery system needs to correct both peripheral tissues and the central nervous system (CNS) because progressive brain pathology occurs in most of these diseases, often accompanied by visceral disease. Intravenous delivery of viral vectors could ultimately prove to be the optimal way of broadly distributing corrective gene therapy to all areas of the brain, as well as transducing peripheral organs. Elegant proof-of-concept studies in several disease models in mice have demonstrated the feasibility of intravenous gene delivery for treating the CNS, peripheral nervous system (PNS) and non-neuronal tissues<sup>39,40</sup>. However, ongoing or planned clinical trials all use, or will use, more-invasive approaches with direct delivery of vectors to the brain. Should efficacy result, less-invasive delivery methods will no doubt be explored in the future<sup>41</sup>.

The next step in the pathogenic cascade, downstream of the genetic defect, is the defective protein. LSDs caused by enzyme deficiencies can be amenable to enzyme replacement therapy (ERT) using recombinant forms of wild-type enzymes delivered through frequent intravenous infusions<sup>42</sup>. Gaucher disease is a prototypical sphingolipid storage disease caused by inherited defects in the gene encoding glucocerebrosidase, leading to storage of its substrate GlcCer<sup>43</sup> (Fig. 2). There are three forms of Gaucher disease: type 1 involves peripheral tissues and organs, type 2 is acute neuronopathic and type 3 can be both neuronopathic and involve peripheral tissues. Type 1 Gaucher disease is characterized by hepatosplenomegaly, haematological defects and bone disease<sup>44</sup>. This was the first disease to benefit from ERT, and the therapy was the first to be commercialized, at that time using a non-recombinant purified enzyme<sup>45</sup>. Despite the remarkable efficacy of ERT in treating patients with type 1 Gaucher disease, lysosomal enzymes do not cross the blood brain barrier, and so cannot be used to treat CNS manifestations in patients with type 2 and type 3 disease. The success of ERT in type 1 Gaucher patients catalysed the development of ERT for several other LSDs, including another sphingolipid storage disease, Fabry disease<sup>23</sup>. Commercially, ERT showed that the development of therapies for rare diseases could be highly profitable, which has catalysed the development of a number of competing ERT products from multiple companies<sup>46</sup>. The high costs of ERT and other treatments for rare diseases limits their global access and raises the spectre of health-economics arguments about cost effectiveness<sup>47</sup>. Owing to low patient numbers, therapies that could be used in multiple LSDs would be attractive. One experimental therapy that has shown broad benefit in *in vitro* models of LSDs is heat shock protein 70 (HSP70), which stabilizes lysosomal membranes and promotes clearance of storage by an incompletely understood mechanism<sup>48,49</sup>. This biologic therapy is currently in preclinical development.

The only other approved therapy for LSDs that involve GSL storage is an oral, small-molecule approach termed substrate reduction therapy (SRT)<sup>50</sup>, of which miglustat is an example. GSL biosynthesis is inhibited by this drug, reducing the number of GSLs requiring lysosomal catabolism, thereby reducing storage<sup>51</sup>. Miglustat is approved for the treatment of type 1 Gaucher disease<sup>52</sup>, as well as for NPC disease. For patients with Gaucher disease, miglustat offers an oral alternative for those who are unable or unwilling to receive intravenous ERT<sup>53</sup>. Miglustat is currently the only specific, approved disease-modifying drug for NPC disease<sup>54</sup>. It crosses the blood-brain barrier and slows progression of CNS disease<sup>55</sup>. Early intervention has the most clinical benefit and prolongs life by reducing aspiration pneumonia, owing to improved swallowing<sup>56</sup>. NPC disease is not caused by an enzyme deficiency, but most commonly by mutations in a multimembrane spanning transporter called NPC1, which is located in the limiting membrane of the lysosome<sup>57</sup>. New strategies to treat NPC will require the downstream steps in the pathogenic cascade (Fig. 3) to be identified and targeted; this disease is not currently a promising candidate for gene therapy because cells are not amenable to cross-correction by transduced cells<sup>58</sup>.



**Figure 2 | Glycosphingolipid catabolism and associated lysosomal storage diseases.** The enzymes that catalyse the catabolism of one metabolite to another are shown in red, and the diseases that result from defects of these enzymes are shown in blue. ASA, arylsulphatase A; GBA,  $\beta$ -glucocerebrosidase; GLA,  $\alpha$ -galactosidase; GLB,  $\beta$ -galactosidase; HEXA,  $\beta$ -hexosaminidase A; HEXB,  $\beta$ -hexosaminidase B; NA, neuraminidase; SMase, acid sphingomyelinase.



**Table 2 | Approved therapies and those in clinical trials for lysosomal storage diseases**

Therapeutic approach	Disease	Status	Approved product or ClinicalTrials.gov identifier
ERT	Gaucher disease*	Approved	Cerezyme
			VPRIV
	Fabry*	Approved	Ellyso
			Fabrazyme
	MPS I	Approved	Replagal
	MPS II	Approved	Aldurazyme
	MPS VI	Approved	Elaprase
	Pompe	Approved	Naglazyme
	NPB	Clinical trials	Myozyme
			NCT02004704 (acid sphingomyelinase)
SRT	Type 1 Gaucher disease*	Approved	Miglustat
	NPC*	Approved (not by FDA)	Miglustat
SRT	Type 1 Gaucher disease*	Ongoing clinical trials	NCT00358150 (eliglustat)
Enzyme enhancement (chaperone therapy)	Fabry*	Ongoing clinical trials	NCT01458119 (migalastat)
	Pompe	Ongoing clinical trials	NCT01380743 (duvoglustat)
Gene therapy	Tay–Sachs*	Trials pending or recruiting	NCT01869270
BMT	Multiple	In clinical practice	N/A

\*Indicates sphingolipid storage diseases. BMT, bone marrow transplantation; ERT, enzyme replacement therapy; MPS, mucopolysaccharidosis; NPB, Niemann–Pick type B; SRT, substrate reduction therapy.

### The benefits and challenges of combination therapy

The concept and application of combination therapy is well established in medicine and involves drugs that target two or more unique steps in a disease process to maximize the clinical benefit for the patient (additive or synergistic benefit). A complex disease such as rheumatoid arthritis is a good example of one in which multiple, independent steps in inflammatory cascades are targeted with small molecules and biologics<sup>59</sup>. In LSDs, in which there is a monogenic cause, targeting the gene or defective protein has logically driven most of the therapeutic efforts for the past two decades<sup>42,58</sup>. However, gene therapy is not yet in routine clinical use for these diseases and, because ERT cannot target the brain, this form of therapy is limited to the management of peripheral organ dysfunction. Several factors are directing new approaches to therapy. We now have much more insight into the pathogenesis of LSDs: for example, we know that dysregulated inflammation is a downstream consequence of lysosomal storage in multiple sphingolipid LSDs<sup>2</sup>. The mechanisms leading to dysregulation of inflammatory signalling remain unclear; however, it can be treated with existing cheap and effective drugs. For example, studies conducted in some animal models have shown that non-steroidal anti-inflammatory drugs (NSAIDs) have modest, but significant, benefits when used as a monotherapy<sup>60,61</sup>. However, synergy has been achieved when anti-inflammatory drugs have been combined with other treatments, including bone-marrow transplantation and SRT<sup>62</sup>. The other factor driving combination-therapy approaches is economic. Specific therapies developed for treating LSDs (for example ERT and SRT) are extremely expensive, so combination therapy that uses multiple high-cost drugs would be unrealistic. There is therefore an urgent need to identify steps in the pathogenic cascade so that they can be targeted with pre-existing, cheaper medicines, the safety profiles of which are already known, thereby allowing rapid translation into the clinic.

### New developments in our understanding of LSDs

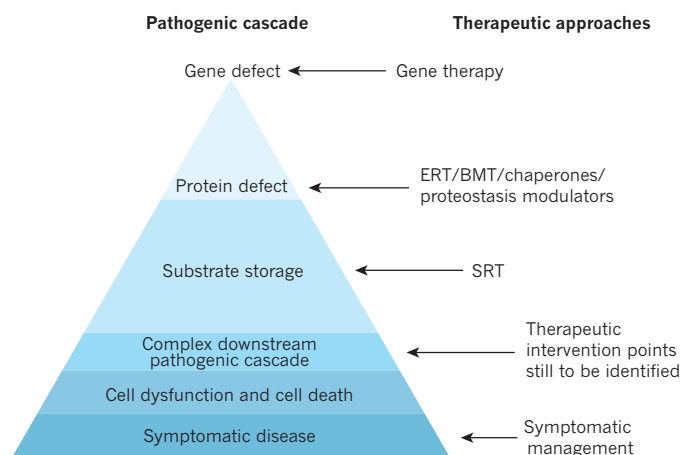
Several important developments are shedding light on aspects of lysosome biology, LSD pathogenesis and new approaches to treatment. The first of these addresses a fundamental question: are lysosomal proteins coordinately regulated? Using a bioinformatics approach, the CLEAR (coordinated lysosomal expression and regulation) network of genes was identified, demonstrating that genes encoding proteins involved in lysosomal function have coordinated transcription controlled by a master regulator, the helix–loop–helix transcription factor TFEB<sup>63–65</sup>. A transcriptional regulator, ZKSCAN3, with opposing activity has also been identified, providing insight into the negative regulation of lysosome biogenesis<sup>66,67</sup>. These key studies demonstrated that lysosome biogenesis and function are tightly regulated and imply that TFEB might have a role in LSDs in which storage triggers expansion of the lysosomal system.

Furthermore, the characterization of this regulated gene network offers new approaches to control lysosomal biogenesis and potentially increase flux through the lysosomal system, which could have therapeutic potential in a number of lysosomal disorders.

Lysosomes, in addition to playing a central part in macromolecule catabolism, also mediate plasma-membrane repair<sup>68,69</sup>. During the process of lysosome–cell-membrane fusion, exosomes are released; TFEB activation has been shown to greatly enhance this process in a mucolipin-1-dependent manner<sup>70</sup>. *MCOLN1* encodes mucolipin-1, which, when defective, causes mucopolipidosis type IV, another LSD. The function of mucolipin-1 is incompletely understood, but has been implicated in the regulation of aspects of acidic store calcium homeostasis, suggesting TFEB requires lysosomal calcium for its activation or translocation<sup>70</sup>.

TFEB activation could therefore offer a practical means to remove storage vesicles by promoting exocytosis. A proof-of-concept study has shown the benefit of this approach in a mouse model of multiple sulphatase deficiency, in which overexpression of TFEB *in vivo* led to reduced glycosaminoglycan storage and reduced inflammation<sup>70</sup>. Drugs that can differentially regulate TFEB expression, and potentially other regulators of lysosome function, could therefore represent a new class of therapeutic agent for treating LSDs in the future.

In the context of sphingolipid LSDs, the question is whether this approach would reduce stored insoluble lipids (in contrast to glycosaminoglycans). This is because the storage burden released by one cell in the form of exosomes might be taken up by macrophages, and potentially other immune cells, which can internalize exosome-sized particles and redistribute the lipid. However, for those diseases with neuropathology, lipid redistribution from the CNS to peripheral tissues could still be of clinical benefit<sup>71</sup>. Interestingly, a new therapeutic approach for the treatment of patients with NPC disease is the use of  $\beta$ -cyclodextrin, based on efficacy in murine and feline models of the disease<sup>72–74</sup>.  $\beta$ -cyclodextrin has the ability to promote exocytosis by signalling plasma-membrane damage in NPC cells *in vitro*<sup>75</sup>. Early intervention with  $\beta$ -cyclodextrin in mice cleared storage lipids (sphingosine, GSLs and cholesterol) from the brain, despite a lack of evidence that it can cross the blood–brain barrier<sup>76</sup>. Peripheral organs such as the liver also had reduced storage levels, except for the lung in which foamy lipid-laden macrophages were found, suggesting lipid redistribution to pulmonary macrophages<sup>71</sup>. When a small number of patients with NPC disease were given  $\beta$ -cyclodextrin in individual investigational new drug studies (in the United States, Spain and Brazil), stored lipids were found to redistribute to circulating B cells, possibly as a result of exosome uptake<sup>77</sup>. Clearly, careful evaluation of the impact of stimulating exocytosis as a therapeutic approach in the sphingolipidoses will be an important step towards understanding its therapeutic potential in LSDs involving different biochemical classes of stored metabolites.



**Figure 3 | The pathogenic cascade and therapeutic approaches.** Defects in individual genes and proteins lead to the storage of substrates. Through a complex downstream pathogenic cascade, the steps of which are still to be determined, this storage leads to dysfunction and death at the cellular level, and to symptomatic disease. Treating disease requires an understanding of these downstream steps and the points at which therapeutic intervention is possible. Only bone marrow transplantation (BMT), enzyme replacement therapy (ERT), substrate reduction therapy (SRT) and symptomatic management are approved or in routine clinical practice.

### Unusual mechanisms of cell death in some LSDs

Most LSDs involve progressive neurodegeneration, it is therefore of great importance to understand cell-death pathways, which represent another potential therapeutic target. Several studies in mouse models of LSDs involving ganglioside storage (for example, Sandhoff disease and GM1 gangliosidosis) have provided evidence for a form of programmed cell death called caspase-dependent apoptosis<sup>78,79</sup>.

However, another study<sup>80</sup> has uncovered an unusual cell-death mechanism in two sphingolipid storage diseases, type 2 Gaucher disease and Krabbe disease. These diseases share a common biochemical hallmark, the storage of simple monohexosyl GSLs. Type 2 Gaucher disease is a severe neuronopathic form of the disease, leading to death in infancy; and is refractory to ERT because the disease manifests in the brain.

#### BOX 1

## Necroptosis in Gaucher and Krabbe disease

- Is necroptotic cell death restricted to lipid storage diseases that involve storage of monohexosyl ceramides?
- If necroptotic cell death is selective for monohexosyl ceramides, what steps in the necroptosis pathway do these lipids target?
- Where in the cell do storage lipids intersect with the necroptotic cell-death pathway?
- What level of glucosylceramide (GlcCer) or galactosylceramide (GalCer) is required to trigger necroptotic cell death?
- Why is necroptosis not a feature of Niemann–Pick type C (NPC) with low levels of GlcCer storage? Is it because GlcCer levels are lower or because there is a lack of GlcCer transfer to other organelles or membranes in NPC disease cells?
- Is necroptotic death involved in other neurodegenerative diseases (such as Parkinson's disease) in which the mechanism of cell death is still unclear or controversial?
- How easy will it be to generate selective inhibitors of RIPK3 to test in animal models and potentially in the clinic?

Type 2 Gaucher disease results from a profound loss of glucocerebrosidase activity (owing to mutations in *GBA*) that leads to storage of GlcCer and its lyso-derivative glucosylsphingosine. Krabbe disease is another severe neurodegenerative LSD caused by storage of a different monohexosylceramide, the galacto-glycoconjugate GalCer, and its lyso-derivative psychosine<sup>81,82</sup>.

In addition to apoptosis, which is perhaps the most intensively researched form of programmed cell death<sup>83</sup>, necroptosis<sup>84,85</sup> is also a major programmed cell-death pathway. This pathway is dependent on the kinases RIPK1 (ref. 86) and RIPK3, whereby RIPK3 (ref. 87) mediates caspase-independent cell death. Futerman and colleagues found that in the brains of mice that were deficient for Gba in neural and glial progenitor cells (a model of acute neuronopathic Gaucher disease) there was an absence of TUNEL positive cells, which are indicative of apoptotic cell death, even at time points when neurodegeneration was advanced, indicating an absence of caspase activity. Furthermore, caspase expression was normal throughout the disease course, supporting a caspase-independent mechanism of neuronal cell death. Levels of RIPK1 and RIPK3 were found to be significantly elevated in these mice, and, in the same study, RIPK1 was also found to be increased in the post-mortem brain of a patient with type 2 Gaucher disease. In the mice, RIPK3 was elevated in both microglia and in neurons, suggesting a contribution to neuroinflammation and neuronal cell death, respectively. Furthermore, the levels of RIPK3 correlated with the extent of pathology in different regions of the brain. The necroptosis pathway was not activated in mouse models of other LSDs, including NPC disease, Sandhoff disease and GM1 gangliosidosis. However, it was activated in Krabbe disease.

The authors also tested the therapeutic potential of targeting necroptosis by pharmacologically inducing neuronopathic Gaucher disease (using conduritol B epoxide, CBE; an irreversible inhibitor of glucocerebrosidase) in *Ripk3*-deficient mice. *Ripk3* heterozygous mice survived for 20–30 days when injected with CBE, whereas the lifespan of *Ripk3*-deficient mice treated with CBE was extended to 100–160 days, thereby demonstrating the therapeutic benefit of targeting RIPK3 activity *in vivo*. The development of pharmacological inhibitors of RIPK3 activity should facilitate clinical investigations of this novel therapeutic approach for these two currently intractable LSDs. Answering some of the questions raised by this study (Box 1) will help us to understand how some storage lipids lead to necroptosis and to define the range of diseases that could benefit from treatment with RIPK3 inhibitors.

### Effects of lipid storage beyond the lysosome

The study of necroptotic pathways in the GlcCer and GalCer storage diseases discussed prompts another question, namely, where in the cell do the storage lipids affect cellular function, leading to pathology? In an *Npc1* mouse model, no elevation of *Ripk1* or *Ripk3* was observed in the brain<sup>80</sup>. This is surprising because GlcCer is one of the GSLs stored in NPC, albeit at lower levels than in neuronopathic Gaucher disease. In most clinical cases, NPC is caused by inherited defects in the integral lysosomal membrane protein, NPC1. Based on the protein's homology to the bacterial RND permease family, it probably facilitates the efflux of multiple substrates out of the lysosome<sup>57,88,89</sup>. In NPC1-deficient cells, late-endosome–lysosome fusion is blocked<sup>90</sup> as a result of reduced calcium release<sup>28</sup> from the lysosome, so that accumulating lipids are trapped in the late endosomes and cannot move from one membrane to another due to defective trafficking<sup>25</sup>. This suggests that the stored monohexosylceramides that trigger necroptosis in Gaucher and Krabbe disease might do so by acting outside the lysosome. The reasoning behind this speculation is that GlcCer is also stored in NPC disease, but is retained in late endosomes/lysosomes as a result of blocked late-endosome–lysosome fusion. This concept is reminiscent of previous studies that implicated defective calcium homeostasis in LSDs<sup>91</sup>. GlcCer accumulation in Gaucher disease, GM2 ganglioside storage in Tay–Sachs and Sandhoff diseases, and GM1 storage in GM1 gangliosidosis were all found to cause dysregulated endoplasmic reticulum (ER)



calcium homeostasis through unique mechanisms that lead to elevated levels of cytosolic calcium. In GM1 gangliosidosis cells, it was found that stored lipids escape the lysosome and are present in ER membranes, hence their impact on ER calcium homeostasis<sup>78</sup>. These data suggest that there is lipid exchange from the lysosome to other membranes, including the ER. By contrast, NPC disease, in which GM2 ganglioside and GlcCer accumulate, has normal ER calcium homeostasis, presumably as a consequence of the block in late-endosome–lysosome fusion that prevents lipids stored in late endosomes from moving into other membranes<sup>25</sup>. This highlights the significance of not only the biochemical nature of the stored material and its overall levels but also its intracellular distribution. It is conceivable that in several LSDs it is the storage lipids in compartments other than the lysosome that contribute to key aspects of pathogenesis<sup>25</sup>.

### Unexpected links to Parkinson's disease

As more genetic studies have been conducted to elucidate pathways that are involved in neurodegenerative diseases, a common theme has emerged, namely defective endosomal/lysosomal transport is a common, shared feature<sup>92</sup>. The mechanisms that cause neurons to malfunction and ultimately die when the function of late endosomes or lysosomes is defective remain obscure. However, it is clear that they have a greater dependency on correct trafficking and endocytic system function than many other cell types<sup>92</sup>. This could reflect their unique anatomical, metabolic and/or functional characteristics, but at present we lack a good mechanistic understanding of their differential vulnerability. This topic has recently been discussed in detail in an excellent review<sup>92</sup>, as has the dysregulation of autophagy, another common feature of neurodegenerative diseases<sup>93</sup>. If we want proof of the importance of normal lysosomal function for neuronal health, we need look no further than the LSDs, in which defects in many different, unique aspects of late-endosome or lysosome function commonly result in a relentless neurodegenerative clinical course<sup>94</sup>. Historically, LSDs have been studied as discrete metabolic diseases in their own right. However, in the past few years some unanticipated links with more common diseases have emerged, placing the lysosome at centre stage in a range of severe human diseases. Probably the most extensively researched example is the link between Gaucher and Parkinson's disease<sup>95</sup>.

Cases of patients with Gaucher disease who had parkinsonian symptoms were reported in the late 1990s<sup>96–99</sup>. The clinical presentation of parkinsonian symptoms in one study cohort was bilateral akinetic rigidity and poor response to levodopa treatment<sup>100</sup>, but there was variability in clinical presentation and age of onset<sup>101</sup>. Because having a rare monogenic disease is unfortunately not a barrier to developing more common diseases, this could simply have reflected chance. However, over time, more cases were reported<sup>102</sup>, and it was observed<sup>103</sup> that the frequency of Parkinson's disease in unaffected family members of patients with Gaucher disease (some of whom were carriers of the *GBA* gene mutation) was higher than would be expected in the general population. Genetic studies to determine whether there is a significant link between Gaucher and Parkinson's disease showed that carrying a *GBA* mutation was the highest risk factor for developing Parkinson's, elevating risk fivefold for Parkinson's and eightfold for Lewy body dementia, a condition associated with Parkinson's<sup>104,105</sup>. Moreover, heterozygous carriers of *GBA* mutations that are responsible in homozygotes for the development of Gaucher disease seem to be affected by parkinsonism with equal frequency, strongly suggesting the presence of a 'toxic allele' effect. A number of interesting questions arise from these studies of the mechanism linking Gaucher and Parkinson's (Box 2). Anecdotal reports have also described Parkinson's disease in patients with other LSDs<sup>106</sup>. Additional genes involved in lysosomal function have been linked to Parkinson's disease and have been reviewed<sup>107</sup>. These findings suggest that lysosomal dysfunction might be linked to other neurodegenerative diseases. This will no doubt be a hot topic of research in the field

### BOX 2

## The link between Gaucher and Parkinson's disease

- How does carrying one mutant *GBA* allele increase the risk of Parkinson's disease?
- Most mutant *GBA* alleles that are associated with Parkinson's disease lead to misfolded proteins, so is there a gain-of-function mechanism?
- Does mutant glucocerebrosidase enhance  $\alpha$ -synuclein aggregation and, if so, how?
- Does the *GBA* mutant allele impair lysosomal function, overloading other pathways, including the proteasome pathway?
- Does the *GBA* mutation in the heterozygous state lead to altered glucosylceramide metabolism in certain specialized cell types, such as dopaminergic neurons?
- Is it possible that necroptotic cell death in neuronopathic Gaucher disease also occurs in Parkinson's disease?
- Why do most individuals with a mutant *GBA* allele not develop Parkinson's disease, and what are the other risk factors?

for years to come, and I predict that this is the tip of the iceberg in terms of links between lysosomal dysfunction and more-common human diseases.

### Future directions

Remarkable progress has been made in our understanding of classical LSD, which has led to a number of approved therapies that are improving the quality of life for patients. Treating the brain remains an unmet clinical need, and innovative experimental therapies that cross the blood–brain barrier or are delivered directly to the CNS are currently under investigation. There is also a growing appreciation of the benefits of combination therapy and the part it will play in the clinical management of patients in the future. A better understanding of the pathogenic cascade in these diseases must be a priority, because this will no doubt identify adjunctive therapies that target novel clinical intervention points, potentially with approved drugs, thereby speeding the path to translation to the clinic. The unanticipated links between Gaucher and Parkinson's disease are provoking a re-think of lysosomal disorders beyond inborn errors of metabolism. How many other more-common diseases involve lysosomal dysfunction? Answering this question will have therapeutic ramifications because therapies developed to treat rare monogenic diseases might have unanticipated uses in much more common diseases that also involve lysosomal dysfunction. Therefore, paradoxically, lysosomal diseases could be much more common than we currently think. Knowledge gained from studying rare monogenic LSDs could help us to unravel disease mechanisms of sporadic polygenic diseases such as Parkinson's that may also involve lysosomal dysfunction. Understanding disease mechanisms in the GSL metabolic diseases will also provide fundamental insights into the biological roles of this enigmatic family of lipids. ■

Received 24 January; accepted 14 March 2014.

1. Sandhoff, K. & Kolter, T. Biosynthesis and degradation of mammalian glycosphingolipids. *Phil. Trans. R. Soc. Lond. B* **358**, 847–861 (2003).
2. Vitner, E. B., Platt, F. M. & Futerman, A. H. Common and uncommon pathogenic cascades in lysosomal storage diseases. *J. Biol. Chem.* **285**, 20423–20427 (2010).
3. Schnaar, R. L., Suzuki, A. & Stanley, P. in *Essentials of Glycobiology* (eds Varki, A. et al.) (Cold Spring Harbour Laboratory Press, 2009).
4. Schnaar, R. L. in *Neuroglycobiology* (eds Fukuda, M., Rutishauser, U., Schnaar, R. L., & Yamaguchi, Y.) 95–113 (Oxford University Press, 2005).
5. Park, J. W., Park, W. J. & Futerman, A. H. Ceramide synthases as potential targets for therapeutic intervention in human diseases. *Biochim. Biophys. Acta* **1841**, 671–681 (2013).

6. Yamashita, T. *et al.* A vital role for glycosphingolipid synthesis during development and differentiation. *Proc. Natl Acad. Sci. USA* **96**, 9142–9147 (1999).
7. Proia, R. L. Glycosphingolipid functions: insights from engineered mouse models. *Phil. Trans. R. Soc. Lond. B* **358**, 879–883 (2003).
8. Simpson, M. A. *et al.* Infantile-onset symptomatic epilepsy syndrome caused by a homozygous loss-of-function mutation of GM3 synthase. *Nature Genet.* **36**, 1225–1229 (2004).  
**This study was the first to provide a genetically and biochemically proven example of a human GSL biosynthetic disease.**
9. Harlalka, G. V. *et al.* Mutations in B4GALNT1 (GM2 synthase) underlie a new disorder of ganglioside biosynthesis. *Brain* **136**, 3618–3624 (2013).
10. Boukhris, A. *et al.* Alteration of ganglioside biosynthesis responsible for complex hereditary spastic paraplegia. *Am. J. Hum. Genet.* **93**, 118–123 (2013).  
**References 9 and 10 both report the second genetically proven example of a human GSL biosynthetic disease.**
11. Proia, R. L. Gangliosides help stabilize the brain. *Nature Genet.* **36**, 1147–1148 (2004).
12. Schultz, M. L., Tecedor, L., Chang, M. & Davidson, B. L. Clarifying lysosomal storage diseases. *Trends Neurosci.* **34**, 401–410 (2011).
13. Wraith, J. E. in *Lysosomal Disorders of the Brain* (eds Platt, F. M. & Walkley, S. U.) 50–77 (Oxford University Press, 2004).
14. Rapola, J. Lysosomal storage diseases in adults. *Pathol. Res. Pract.* **190**, 759–766 (1994).
15. Sedel, F. Niemann-Pick diseases in adults. *Rev. Med. Interne* **28** (suppl. 4), S292–S293 (2007).
16. van der Beek, N. A. *et al.* Clinical features and predictors for disease natural progression in adults with Pompe disease: a nationwide prospective observational study. *Orphanet J. Rare Dis.* **7**, 88 (2012).
17. Hopwood, J. J., Crawley, A. C. & Taylor, R. M. in *Lysosomal disorders of the brain* Vol. 1 (eds Platt, F. M. & Walkley, S. U.) 257–289 (Oxford University Press, 2004).
18. Hemsley, K. M. & Hopwood, J. J. Lessons learnt from animal models: pathophysiology of neuropathic lysosomal storage disorders. *J. Inher. Metab. Dis.* **33**, 363–371 (2010).
19. Zeng, B. J. *et al.* Spontaneous appearance of Tay-Sachs disease in an animal model. *Mol. Genet. Metab.* **95**, 59–65 (2008).
20. Zervas, M., Somers, K. L., Thrall, M. A. & Walkley, S. U. Critical role for glycosphingolipids in Niemann-Pick disease type C. *Curr. Biol.* **11**, 1283–1287 (2001).  
**The study reports the first animal model evidence that miglustat might be a disease-modifying treatment for NPC disease.**
21. Patterson, M. C., Vecchio, D., Prady, H., Abel, L. & Wraith, J. E. Miglustat for treatment of Niemann-Pick C disease: a randomised controlled study. *Lancet Neurol.* **6**, 765–772 (2007).  
**This investigation confirmed that miglustat is a disease-modifying treatment for patients with NPC disease.**
22. Ioannou, Y. A., Zeidner, K. M., Gordon, R. E. & Desnick, R. J. Fabry disease: preclinical studies demonstrate the effectiveness of alpha-galactosidase A replacement in enzyme-deficient mice. *Am. J. Hum. Genet.* **68**, 14–25 (2001).
23. Brady, R. O., Murray, G. J., Moore, D. F. & Schiffmann, R. Enzyme replacement therapy in Fabry disease. *J. Inher. Metab. Dis.* **24**, 18–24, discussion 11–12 (2001).
24. Laurijssens, B., Aujard, F. & Rahman, A. Animal models of Alzheimer's disease and drug development. *Drug Discov. Today Technol.* **10**, e319–e327 (2013).
25. Platt, F. M., Boland, B. & van der Spoel, A. C. The cell biology of disease: lysosomal storage disorders: the cellular impact of lysosomal dysfunction. *J. Cell Biol.* **199**, 723–734 (2012).
26. Saftig, P. & Klumperman, J. Lysosome biogenesis and lysosomal membrane proteins: trafficking meets function. *Nature Rev. Mol. Cell Biol.* **10**, 623–635 (2009).
27. Rosenbaum, A. I. & Maxfield, F. R. Niemann-Pick type C disease: molecular mechanisms and potential therapeutic approaches. *J. Neurochem.* **116**, 789–795 (2011).
28. Lloyd-Evans, E. *et al.* Niemann-Pick disease type C1 is a sphingosine storage disease that causes deregulation of lysosomal calcium. *Nature Med.* **14**, 1247–1255 (2008).
29. Lloyd-Evans, E. & Platt, F. M. Lipids on trial: the search for the offending metabolite in Niemann-Pick type C disease. *Traffic* **11**, 419–428 (2010).
30. Patterson, M. C. *et al.* Recommendations for the diagnosis and management of Niemann-Pick disease type C: an update. *Mol. Genet. Metab.* **106**, 330–344 (2012).
31. Tomanin, R. *et al.* Gene therapy approaches for lysosomal storage disorders, a good model for the treatment of mendelian diseases. *Acta Paediatr.* **101**, 692–701 (2012).
32. Cachón-González, M. B. *et al.* Effective gene therapy in an authentic model of Tay-Sachs-related diseases. *Proc. Natl Acad. Sci. USA* **103**, 10373–10378 (2006).
33. Ellinwood, N. M. *et al.* Safe, efficient, and reproducible gene therapy of the brain in the dog models of Sanfilippo and Hurler syndromes. *Mol. Ther.* **19**, 251–259 (2011).
34. Ziegler, R. J. *et al.* Correction of the nonlinear dose response improves the viability of adenoviral vectors for gene therapy of fabry disease. *Hum. Gene Ther.* **13**, 935–945 (2002).
35. Yew, N. S. & Cheng, S. H. Gene therapy for lysosomal storage disorders. *Pediatr. Endocrinol. Rev.* **11** (suppl. 1), 99–109 (2013).
36. Bradbury, A. M. *et al.* Therapeutic response in feline Sandhoff disease despite immunity to intracranial gene therapy. *Mol. Ther.* **21**, 1306–1315 (2013).
37. Coutelle, C. & Waddington, S. N. The concept of prenatal gene therapy. *Methods Mol. Biol.* **891**, 1–7 (2012).
38. Mattar, C. N. *et al.* The case for intrauterine gene therapy. *Best Pract. Res. Clin. Obstet. Gynaecol.* **26**, 697–709 (2012).
39. Rahim, A. A. *et al.* Intravenous administration of AAV2/9 to the fetal and neonatal mouse leads to differential targeting of CNS cell types and extensive transduction of the nervous system. *FASEB J.* **25**, 3505–3518 (2011).
40. Foust, K. D. *et al.* Rescue of the spinal muscular atrophy phenotype in a mouse model by early postnatal delivery of SMN. *Nature Biotechnol.* **28**, 271–274 (2010).
41. Nagabhushan Kalburgi, S., Khan, N. N. & Gray, S. J. Recent gene therapy advancements for neurological diseases. *Discov. Med.* **15**, 111–119 (2013).
42. Brady, R. O. Enzyme replacement for lysosomal diseases. *Annu. Rev. Med.* **57**, 283–296 (2006).  
**This review describes the pioneering development of the first ERTs for LSDs.**
43. Jmoudiak, M. & Futerman, A. H. Gaucher disease: pathological mechanisms and modern management. *Br. J. Haematol.* **129**, 178–188 (2005).
44. Cox, T. M. & Schofield, J. P. Gaucher's disease: clinical features and natural history. *Baillieres Clin. Haematol.* **10**, 657–689 (1997).
45. Brady, R. O. Enzyme replacement therapy: conception, chaos and culmination. *Phil. Trans. R. Soc. Lond. B* **358**, 915–919 (2003).
46. Cox, T. M. Competing for the treasure in exceptions. *Am. J. Hematol.* **88**, 163–165 (2013).
47. Rombach, S. M., Hollak, C. E., Linthorst, G. E. & Dijkgraaf, M. G. Cost-effectiveness of enzyme replacement therapy for Fabry disease. *Orphanet J. Rare Dis.* **8**, 29 (2013).
48. Kirkegaard, T. *et al.* Hsp70 stabilizes lysosomes and reverts Niemann-Pick disease-associated lysosomal pathology. *Nature* **463**, 549–553 (2010).
49. Petersen, N. H. & Kirkegaard, T. HSP70 and lysosomal storage disorders: novel therapeutic opportunities. *Biochem. Soc. Trans.* **38**, 1479–1483 (2010).
50. Platt, F. M. & Jeyakumar, M. Substrate reduction therapy. *Acta Paediatr. Suppl.* **97**, 88–93 (2008).
51. Platt, F. M., Neises, G. R., Dwek, R. A. & Butters, T. D. N-butyldeoxynojirimycin is a novel inhibitor of glycolipid biosynthesis. *J. Biol. Chem.* **269**, 8362–8365 (1994).
52. Cox, T. *et al.* Novel oral treatment of Gaucher's disease with N-butyldeoxynojirimycin (OGT 918) to decrease substrate biosynthesis. *Lancet* **355**, 1481–1485 (2000).
53. Lachmann, R. H. Miglustat: substrate reduction therapy for glycosphingolipid lysosomal storage disorders. *Drugs Today (Barc.)* **42**, 29–38 (2006).
54. Lyseng-Williamson, K. A. Miglustat: a review of its use in Niemann-Pick disease type C. *Drugs* **74**, 61–74 (2014).
55. Chien, Y. H. *et al.* Long-term efficacy of miglustat in paediatric patients with Niemann-Pick disease type C. *J. Inher. Metab. Dis.* **36**, 129–137 (2012).
56. Walterfang, M. *et al.* Dysphagia as a risk factor for mortality in Niemann-Pick disease type C: systematic literature review and evidence from studies with miglustat. *Orphanet J. Rare Dis.* **7**, 76 (2012).
57. Ioannou, Y. A. The structure and function of the Niemann-Pick C1 protein. *Mol. Genet. Metab.* **71**, 175–181 (2000).
58. Neufeld, E. F. From serendipity to therapy. *Annu. Rev. Biochem.* **80**, 1–15 (2011).  
**This is a review of the pioneering discovery of cross-correction by lysosomal enzymes that formed the basis for ERT.**
59. Ruderman, E. M. The role of concomitant methotrexate in biologic therapy for rheumatoid arthritis. *Bull. Hosp. Jt. Dis.* **71** (suppl. 1), S29–S32 (2013).
60. Smith, D., Wallom, K. L., Williams, I. M., Jeyakumar, M. & Platt, F. M. Beneficial effects of anti-inflammatory therapy in a mouse model of Niemann-Pick disease type C1. *Neurobiol. Dis.* **36**, 242–251 (2009).
61. Jeyakumar, M. *et al.* NSAIDs increase survival in the Sandhoff disease mouse: Synergy with N-butyldeoxynojirimycin. *Ann. Neurol.* **56**, 642–649 (2004).
62. Jeyakumar, M. *et al.* Enhanced survival in Sandhoff disease mice receiving a combination of substrate deprivation therapy and bone marrow transplantation. *Blood* **97**, 327–329 (2001).
63. Palmieri, M. *et al.* Characterization of the CLEAR network reveals an integrated control of cellular clearance pathways. *Hum. Mol. Genet.* **20**, 3852–3866 (2011).
64. Sardiello, M. & Ballabio, A. Lysosomal enhancement: a CLEAR answer to cellular degradative needs. *Cell Cycle* **8**, 4021–4022 (2009).
65. Sardiello, M. *et al.* A gene network regulating lysosomal biogenesis and function. *Science* **325**, 473–477 (2009).  
**This article reports the discovery of the key role that the transcription factor TFEB has in regulating lysosome biogenesis.**
66. Chauhan, S. *et al.* ZKSCAN3 is a master transcriptional repressor of autophagy. *Mol. Cell* **50**, 16–28 (2013).
67. Füllgrabe, J., Klionsky, D. J. & Joseph, B. The return of the nucleus: transcriptional and epigenetic control of autophagy. *Nature Rev. Mol. Cell Biol.* **15**, 65–74 (2014).
68. Reddy, A., Caler, E. V. & Andrews, N. W. Plasma membrane repair is mediated by Ca<sup>2+</sup>-regulated exocytosis of lysosomes. *Cell* **106**, 157–169 (2001).  
**The authors of this paper made the discovery of an unanticipated role of lysosomes in plasma-membrane repair.**
69. Jaiswal, J. K., Andrews, N. W. & Simon, S. M. Membrane proximal lysosomes are the major vesicles responsible for calcium-dependent exocytosis in nonsecretory cells. *J. Cell Biol.* **159**, 625–635 (2002).
70. Medina, D. L. *et al.* Transcriptional activation of lysosomal exocytosis promotes cellular clearance. *Dev. Cell* **21**, 421–430 (2011).



71. Liu, B. *et al.* Cyclodextrin overcomes the transport defect in nearly every organ of NPC1 mice leading to excretion of sequestered cholesterol as bile acid. *J. Lipid Res.* **51**, 933–944 (2010).
72. Ramirez, C. M. *et al.* Weekly cyclodextrin administration normalizes cholesterol metabolism in nearly every organ of the Niemann-Pick type C1 mouse and markedly prolongs life. *Pediatr. Res.* **68**, 309–315 (2010).
73. Davidson, C. D. *et al.* Chronic cyclodextrin treatment of murine Niemann-Pick C disease ameliorates neuronal cholesterol and glycosphingolipid storage and disease progression. *PLoS ONE* **4**, e6951 (2009).
74. Stein, V. M. *et al.* Miglustat improves purkinje cell survival and alters microglial phenotype in feline Niemann-Pick disease type C. *J. Neuropathol. Exp. Neurol.* **71**, 434–448 (2012).
75. Chen, F. W., Li, C. & Ioannou, Y. A. Cyclodextrin induces calcium-dependent lysosomal exocytosis. *PLoS ONE* **5**, e15054 (2010).
76. Pontikis, C. C., Davidson, C. D., Walkley, S. U., Platt, F. M. & Begley, D. J. Cyclodextrin alleviates neuronal storage of cholesterol in Niemann-Pick C disease without evidence of detectable blood-brain barrier permeability. *J. Inher. Metab. Dis.* **36**, 491–498 (2013).
77. te Vrugte, D. *et al.* Relative acidic compartment volume as a lysosomal storage disorder-associated biomarker. *J. Clin. Invest.* <http://dx.doi.org/10.1172/JCI72835> (2014).
78. Sano, R. *et al.* GM1-ganglioside accumulation at the mitochondria-associated ER membranes links ER stress to Ca<sup>2+</sup>-dependent mitochondrial apoptosis. *Mol. Cell* **36**, 500–511 (2009).
79. Wu, Y. P. & Proia, R. L. Deletion of macrophage-inflammatory protein 1 alpha retards neurodegeneration in Sandhoff disease mice. *Proc. Natl Acad. Sci. USA* **101**, 8425–8430 (2004).
80. Vitner, E. B. *et al.* RIPK3 as a potential therapeutic target for Gaucher's disease. *Nature Med.* **20**, 204–208 (2014).
- This article reports the discovery of the involvement of the necroptotic cell-death pathway in Gaucher and Krabbe disease.**
81. Wenger, D. A., Rafi, M. A. & Luzi, P. Molecular genetics of Krabbe disease (globoid cell leukodystrophy): diagnostic and clinical implications. *Hum. Mutat.* **10**, 268–279 (1997).
82. Jesionek-Kupnicka, D. *et al.* Krabbe disease: an ultrastructural study of globoid cells and reactive astrocytes at the brain and optic nerves. *Folia Neuropathol.* **35**, 155–162 (1997).
83. Ghavami, S. *et al.* Autophagy and apoptosis dysfunction in neurodegenerative disorders. *Prog. Neurobiol.* **112**, 24–49 (2014).
84. Linkermann, A. & Green, D. R. Necroptosis. *N. Engl. J. Med.* **370**, 455–465 (2014).
85. Christofferson, D. E. & Yuan, J. Necroptosis as an alternative form of programmed cell death. *Curr. Opin. Cell Biol.* **22**, 263–268 (2010).
86. Holler, N. *et al.* Fas triggers an alternative, caspase-8-independent cell death pathway using the kinase RIP as effector molecule. *Nature Immunol.* **1**, 489–495 (2000).
87. Oberst, A. *et al.* Catalytic activity of the caspase-8-FLIP(L) complex inhibits RIPK3-dependent necrosis. *Nature* **471**, 363–367 (2011).
88. Davies, J. P., Chen, F. W. & Ioannou, Y. A. Transmembrane molecular pump activity of Niemann-Pick C1 protein. *Science* **290**, 2295–2298 (2000).
89. Ioannou, Y. A. Multidrug permeases and subcellular cholesterol transport. *Nature Rev. Mol. Cell Biol.* **2**, 657–668 (2001).
90. Liscum, L. Niemann-Pick type C mutations cause lipid traffic jam. *Traffic* **1**, 218–225 (2000).
91. Futerman, A. H. Calcium homeostasis in lysosomal storage diseases. *Int. J. Clin. Pharmacol. Ther.* **48**, S6–S7 (2010).
92. Neefjes, J. & van der Kant, R. Stuck in traffic: an emerging theme in diseases of the nervous system. *Trends Neurosci.* **37**, 66–76 (2014).
93. Nixon, R. A. The role of autophagy in neurodegenerative disease. *Nature Med.* **19**, 983–987 (2013).
94. Wraith, J. E. Lysosomal disorders. *Semin. Neonatol.* **7**, 75–83 (2002).
95. Sidransky, E. Gaucher disease: insights from a rare Mendelian disorder. *Discov. Med.* **14**, 273–281 (2012).
96. Grewal, R. P. *et al.* Neurologic complications of nonneuronopathic Gaucher's disease. *Arch. Neurol.* **48**, 1271–1272 (1991).
97. Neudorfer, O. *et al.* Occurrence of Parkinson's syndrome in type I Gaucher disease. *QJM* **89**, 691–694 (1996).
- The authors of this article were the first to link Gaucher disease and Parkinson's disease.**
98. Machaczka, M., Rucinska, M., Skotnicki, A. B. & Jurczak, W. Parkinson's syndrome preceding clinical manifestation of Gaucher's disease. *Am. J. Hematol.* **61**, 216–217 (1999).
99. Tayebi, N. *et al.* Gaucher disease and parkinsonism: a phenotypic and genotypic characterization. *Mol. Genet. Metab.* **73**, 313–321 (2001).
100. Bembí, B. *et al.* Gaucher's disease with Parkinson's disease: clinical and pathological aspects. *Neurology* **61**, 99–101 (2003).
101. Hruska, K. S., Goker-Alpan, O. & Sidransky, E. Gaucher disease and the synucleinopathies. *J. Biomed. Biotechnol.* **2006**, 78549 (2006).
102. Várkonyi, J. *et al.* Gaucher disease associated with parkinsonism: four further case reports. *Am. J. Med. Genet. A* **116A**, 348–351 (2003).
103. Tayebi, N. *et al.* Gaucher disease with parkinsonian manifestations: does glucocerebrosidase deficiency contribute to a vulnerability to parkinsonism? *Mol. Genet. Metab.* **79**, 104–109 (2003).
104. Sidransky, E. *et al.* Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N. Engl. J. Med.* **361**, 1651–1661 (2009).
- This article confirmed a genetic link between GBA mutations and Parkinson's disease.**
105. Goker-Alpan, O. *et al.* Glucocerebrosidase mutations are an important risk factor for Lewy body disorders. *Neurology* **67**, 908–910 (2006).
106. Shachar, T. *et al.* Lysosomal storage disorders and Parkinson's disease: Gaucher disease and beyond. *Mov. Disord.* **26**, 1593–1604 (2011).
107. Dehay, B. *et al.* Lysosomal impairment in Parkinson's disease. *Mov. Disord.* **28**, 725–732 (2013).
108. Beutler, E. Gaucher disease: multiple lessons from a single gene disorder. *Acta Paediatr. Suppl.* **95**, 103–109 (2006).
109. de Fost, M., Aerts, J. M. & Hollak, C. E. Gaucher disease: from fundamental research to effective therapeutic interventions. *Neth. J. Med.* **61**, 3–8 (2003).
110. Pastores, G. M. & Lien, Y. H. Biochemical and molecular genetic basis of Fabry disease. *J. Am. Soc. Nephrol.* **13** (suppl. 2), S130–S133 (2002).
111. Bersano, A. *et al.* Neurological features of Fabry disease: clinical, pathophysiological aspects and therapy. *Acta Neurol. Scand.* **126**, 77–97 (2012).
112. Brady, R. O. Tay-Sachs disease: the search for the enzymatic defect. *Adv. Genet.* **44**, 51–60 (2001).
113. Mahuran, D. J., Triggs-Raine, B. L., Feigenbaum, A. J. & Gravel, R. A. The molecular basis of Tay-Sachs disease: mutation identification and diagnosis. *Clin. Biochem.* **23**, 409–415 (1990).
114. Shapiro, B. E., Logigian, E. L., Kolodny, E. H. & Pastores, G. M. Late-onset Tay-Sachs disease: the spectrum of peripheral neuropathy in 30 affected patients. *Muscle Nerve* **38**, 1012–1015 (2008).
114. Gravel, R. A. *et al.* in *The Metabolic and Molecular Bases of Inherited Disease* Vol. 3 (eds Scriver, C. R., Beaudet, A. L., Valle, D. & Sly, W. S.) 3827–3876 (McGraw Hill, 2001).
116. Hadfield, M. G., Mamunes, P. & David, R. B. The pathology of Sandhoff's disease. *J. Pathol.* **123**, 137–144 (1977).
117. Suzuki, Y., Sakuraba, H. & Oshima, M. in *The Metabolic and Molecular Bases of Inherited Diseases* Vol. 2 (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 2785–2824 (McGraw Hill, 1995).
118. Suzuki, Y., Oshima, A. & Nanba, E. in *The Metabolic and Molecular Bases of Inherited Disease* Vol. 3 (eds Scriver, C. R., Beaudet, A. L., Valle, D. & Sly, W. S.) 3775–3809 (McGraw Hill, 2001).
119. Yoshida, K., Ikeda, S., Kawaguchi, K. & Yanagisawa, N. Adult GM1 gangliosidosis: immunohistochemical and ultrastructural findings in an autopsy case. *Neurology* **44**, 2376–2382 (1994).
120. Vanier, M. T. Niemann-Pick disease type C. *Orphanet. J. Rare Dis.* **5**, 16 (2010).
121. Mengel, E. *et al.* Niemann-Pick disease type C symptomatology: an expert-based clinical description. *Orphanet. J. Rare Dis.* **8**, 166 (2013).

**Acknowledgements** Thanks to D. Priestman for creating Fig. 1 and to N. Platt for his comments on the manuscript. F.M.P. is a Royal Society Wolfson Research Merit Award holder.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The author declares competing financial interests see [go.nature.com/2i9a53](http://go.nature.com/2i9a53) for details. Readers are welcome to comment on the online version of this paper at [go.nature.com/2i9a53](http://go.nature.com/2i9a53). Correspondence should be addressed to F.M.P. ([frances.platt@pharm.ox.ac.uk](mailto:frances.platt@pharm.ox.ac.uk)).

# The different shades of fat

Vivian Peirce<sup>1</sup>, Stefania Carobbio<sup>1,2</sup> & Antonio Vidal-Puig<sup>1,2</sup>

**Our understanding of adipose tissue biology has progressed rapidly since the turn of the century. White adipose tissue has emerged as a key determinant of healthy metabolism and metabolic dysfunction. This realization is paralleled only by the confirmation that adult humans have heat-dissipating brown adipose tissue, an important contributor to energy balance and a possible therapeutic target for the treatment of metabolic disease. We propose that the development of successful strategies to target brown and white adipose tissues will depend on investigations that elucidate their developmental origins and cell-type-specific functional regulators.**

In the face of the modern obesity epidemic the previously prevailing view that white adipose tissue (WAT) is a passive storage organ for excess calories has been revised. WAT does have an important role in buffering nutrient availability and demand by storing excess calories and preventing the toxic accumulation of excess nutrients in non-adipose tissues<sup>1</sup>; however, WAT also communicates with metabolically relevant organs by secreting so-called adipokines as part of a dynamic endocrine system that regulates nutrient partitioning into peripheral tissues<sup>2</sup>.

The recent confirmation that adult humans have brown adipose tissue (BAT) has transformed our understanding of how adipose tissue regulates metabolism and energy balance once again<sup>3–8</sup>. The primary function of BAT is to maintain core body temperature in response to cold stress by generating heat, a process known as thermogenesis<sup>9</sup>. Like the unilocular white adipocytes in WAT, the multilocular brown adipocytes in BAT also accumulate and store lipids<sup>6</sup>. However, brown adipocytes are distinct from white adipocytes in that their more abundant mitochondria are enriched with uncoupling protein 1 (UCP1), which uncouples substrate oxidation from ATP production so that heat is produced<sup>9</sup>.

The fact that heat production by BAT is an extremely energy-expensive process that burns nutrients has two important caveats. First, to avoid hyperthermia and to regulate energy usage by BAT, thermogenesis in this tissue must be tightly controlled so that heat is specifically produced in response to thermogenic stimuli. Second, when stimulated to produce heat, the so-called activated BAT has an enormous impact on energy balance, at least in rodents. For example, mice housed at room temperature (20–22 °C) would need to consume 60% more food to maintain their body weight compared with mice housed at thermoneutral temperatures (30 °C) to compensate for the increased energy expenditure as a result of increased thermogenesis<sup>10</sup>. Conversely, even at thermoneutrality, BAT ablation or dysfunction in rodents decreases energy expenditure, causing an obese phenotype<sup>11,12</sup>. In addition to its thermogenic effects, activated BAT has a large capacity for glucose and lipid uptake per gram of tissue, as demonstrated by its ability to normalize hyperglycaemia and hyperlipidaemia in mouse models of diabetes and dyslipidaemia<sup>13,14</sup>. In line with its remarkable capacity for substrate oxidation, BAT is activated in rodents in response to excess nutrient consumption, such as eating a high-fat diet, a process known as diet-induced thermogenesis<sup>15</sup>. Therefore, although WAT far outstrips BAT in terms of percentage body mass, activated BAT is an important contributor to nutrient partitioning and utilization, and body weight regulation.

The revelation that adult humans possess such a potent metabolic tissue was accompanied by the discovery that UCP1-expressing thermogenic adipocytes can also be found in WAT in the form of beige adipocytes<sup>16,17</sup>.

In rodents, chronic cold exposure increases thermogenic capacity by increasing the number of brown adipocytes, thereby recruiting BAT mass<sup>9</sup>. As part of the same response, so-called beige or 'brite' (brown in white) cells are also recruited in WAT, resulting in WAT 'browning'<sup>16,17</sup>. Although there is significant overlap between beige and brown adipocytes in their expression of *UCP1* and the genes required for thermogenesis, beige adipocytes have distinctive gene expression signatures, indicating that they are a distinct type of thermogenic cell<sup>18–21</sup>. Unstimulated rodent beige adipocytes are similar to rodent white adipocytes in terms of oxygen consumption rates and uncoupled respiration *in vitro*<sup>18</sup>. Compared with brown adipocytes, however, beige adipocytes have a greater capacity for thermogenic stimulation, with greater inducibility of *UCP1* expression and proportionally increased uncoupled respiration<sup>18</sup>. *UCP1* expression and uncoupled respiration in stimulated beige adipocytes could equal that of stimulated brown adipocytes<sup>18,22</sup>. In summary, when considering adipose tissue involvement in the regulation of energy balance, the contribution of three different 'shades' of fat — white, brown and beige — should be considered.

The discovery of these different shades of fat has important implications when considering the pathogenesis of, and treatments for, the obesity-related metabolic complications that characterize metabolic syndrome. WAT dysfunction is a key pathological phenomenon contributing to metabolic syndrome. Although WAT is able to functionally adapt and expand in response to chronic calorie overconsumption, individuals seem to have a set functional limit beyond which WAT fails to function properly as a storage and endocrine organ<sup>23</sup>. This can result in the lipotoxic accumulation of lipids in peripheral tissues that contributes to insulin resistance and dyslipidaemia<sup>23</sup>. Conceptually, this lipotoxicity could be ameliorated either by increasing the capacity of WAT for nutrient storage and/or by eliminating these excess calories by increasing the activity of BAT. In our opinion, the development of successful adipose-tissue-based therapeutic strategies to treat metabolic syndrome is reliant on a good understanding of basic adipose-tissue biology. We will discuss how rodent models have deepened our understanding of the developmental origins and cell-type-specific functional regulators of white, brown and beige adipocytes, and the current big challenge to translate this information from rodents to humans.

## The origins of adipocytes in rodents

Following their establishment in early life, adipose-tissue depots undergo active remodelling throughout adulthood. Both WAT and BAT have the capacity to expand in response to calorie excess and chronic cold stress, respectively. To enable this expansion, it is important that precursor

<sup>1</sup>University of Cambridge Metabolic Research Laboratories, Level 4, Wellcome Trust-MRC Institute of Metabolic Science, Box 289, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. <sup>2</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK.



cells with adipogenic potential are retained in adulthood<sup>24,25</sup>. We believe that understanding how the lineage origins of these adipocyte precursors biases their potential to adopt a brown-, beige- or white-cell fate on differentiation (reviewed in ref. 26) is essential for the development of therapeutic approaches that target these cells.

### Embryonic and adult origins of adipocytes

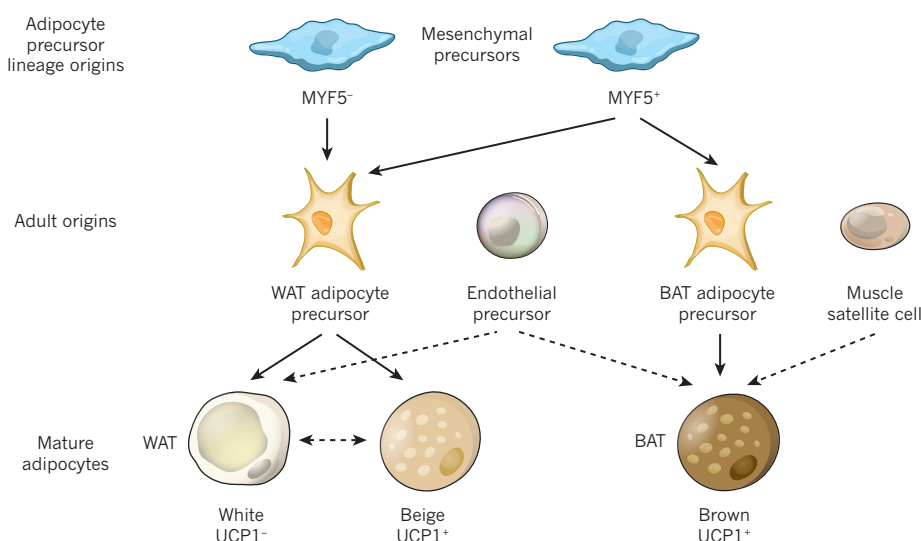
Lineage-tracing studies have shown that both brown adipocytes and myocytes derive from MYF5<sup>+</sup> and PAX7<sup>+</sup> progenitors that originate in the paraxial mesoderm, and that brown adipocytes can also derive from stem-cell-like skeletal muscle satellite cells in adulthood<sup>27–30</sup> (Fig. 1). Moreover, brown adipocyte precursors express muscle-specific transcription factors, and it is known that expression of some myogenic microRNAs (miRNAs) persists in mature brown adipocytes<sup>31,32</sup>. Given the absence of this myogenic gene expression signature in white adipocytes and their precursors, it was expected that white adipocytes would derive preferentially from a MYF5<sup>−</sup> lineage<sup>30–32</sup>. However, this view was recently challenged when the conditional depletion of PTEN driven by MYF5-Cre caused overgrowth of BAT, but also a paradoxical overgrowth of specific WAT depots and a loss of others<sup>33</sup>. Subsequent lineage-tracing studies have demonstrated the presence of some MYF5<sup>+</sup> adipocyte precursors in WAT, indicating that white adipocyte precursors can derive from both MYF5<sup>+</sup> and MYF5<sup>−</sup> lineages<sup>33,34</sup>. Unexpectedly, lineage-tracing studies using the endothelial marker vascular endothelial cadherin or the pre-adipocyte marker *Zfp423* also suggested that some brown and white adipocytes could have endothelial origins, despite a recent study finding that multiple endothelial markers did not label adipocytes<sup>25,35–37</sup>.

Given that beige adipocytes can be derived from white adipocyte precursors *in vitro* by chronic treatment with PPAR- $\gamma$  agonists, it is likely that beige adipocytes share the MYF5<sup>−</sup> lineage origin of most white adipocytes<sup>20,21,38,39</sup>. In terms of adult origins, recent evidence suggests that beige adipocytes can form either directly from white adipocytes or due to the proliferation of precursors<sup>18,40–42</sup>. For example, using mice in which UCP1-expressing cells are labelled both permanently and transiently with fluorescent proteins, beige adipocytes recruited by 7 days of cold exposure were found to arise directly from white adipocytes<sup>40</sup>. Interestingly, these cells return to a white phenotype after warm adaptation, indicating that this interconversion is likely to be bidirectional<sup>40</sup>. However, contrasting results have been found using a mouse model in which

existing adipocytes can be labelled with  $\beta$ -galactosidase ( $\beta$ -gal) using doxycycline, so that any adipocytes generated *de novo* after doxycycline treatment are  $\beta$ -gal<sup>+</sup> (ref. 41). After doxycycline treatment, 3 days of cold exposure induces the appearance of multilocular  $\beta$ -gal<sup>+</sup> adipocytes in subcutaneous WAT, suggesting that these are beige adipocytes produced by clonal expansion of a precursor<sup>41</sup>. This is consistent with results from another study that tracked beige adipogenesis using bromodeoxyuridine accumulation to show that beige adipocytes arise *de novo* in WAT in response to adrenergic stimulation<sup>42</sup>. Subsequent lineage tracing studies indicated that self-renewing PDGFR- $\alpha$ <sup>+</sup> precursors were a significant source of these newly formed beige adipocytes<sup>42</sup>. These are 'bipotential' precursors because they are able to produce both beige and white adipocytes when cultured *in vitro*<sup>42</sup>. In addition to potentially deriving from precursors with both white and beige adipogenic potential, beige adipocytes may derive from dedicated beige adipocyte precursors. In a study characterizing the *in vitro* adipogenic potential of immortalized WAT and BAT precursors, some immortalized WAT precursors differentiated preferentially into beige adipocytes<sup>18</sup>. These results suggest that different types of adipocyte precursors in WAT may vary in their potential to produce beige adipocytes, perhaps due to their lineage origins. For example, PAX3<sup>−</sup> adipocyte precursors and MYF5<sup>−</sup> adipocyte precursors isolated from WAT differentiate into adipocytes with higher expression levels of thermogenic genes compared with PAX3<sup>+</sup> precursors or MYF5<sup>+</sup> precursors, respectively<sup>34,43</sup>. Taken together, these observations suggest that the transcription factor networks that drive adipocyte differentiation towards the brown-, beige- or white-cell fate might be affected by the specific lineage origins of precursor cells.

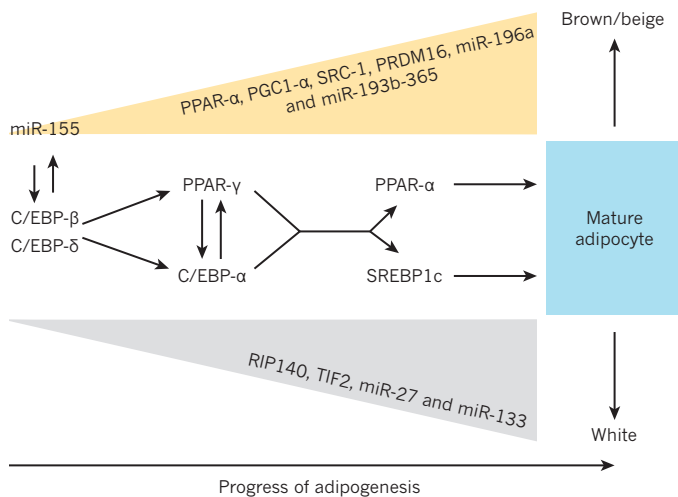
### Key factors driving adipogenesis

Despite the differences in the embryonic origins and physiological functions of brown and white adipocytes, both cell types share a similar transcriptional cascade that controls terminal differentiation into mature adipocytes (reviewed in ref. 44). PPAR- $\gamma$  and three CCAAT/enhancer-binding protein family members (C/EBP- $\alpha$ , C/EBP- $\beta$  and C/EBP- $\delta$ ) have a role in both white and brown adipogenesis by cooperating to promote a transcriptional cascade that leads to and maintains the stable differentiation of the adipocyte<sup>44</sup>. This means that, for example, selective recruitment of thermogenic adipocytes requires the identification of specific factors that robustly drive a thermogenic brown or beige



**Figure 1 | Origins of white, beige and brown adipocytes.** Brown adipose tissue (BAT) is characterized by UCP1<sup>+</sup> brown adipocytes, whereas white adipose tissue (WAT) contains a mixture of UCP1<sup>−</sup> white adipocytes and UCP1<sup>+</sup> beige adipocytes. Adipose tissue expansion in adulthood is achieved mainly by the proliferation and differentiation of adipocyte precursors to produce new adipocytes. WAT and BAT adipocyte precursors derive from mesenchymal precursor cells. Recent evidence

suggests that WAT adipocyte precursors can derive from both MYF5<sup>+</sup> and MYF5<sup>−</sup> lineages, whereas BAT adipocyte precursors derive exclusively from a MYF5<sup>+</sup> lineage. Beige adipocytes can derive from WAT adipocyte precursors and potentially directly from mature white adipocytes. Brown adipocytes can also be generated from stem-cell-like skeletal muscle satellite cells, and both brown and white adipocytes may derive from endothelial precursors.



**Figure 2 | Factors regulating white, and brown or beige adipogenesis.** Adipogenesis is driven by a cascade of transcription factors (C/EBP- $\delta$ , C/EBP- $\beta$ , C/EBP- $\alpha$ , PPAR- $\gamma$ , PGC1- $\alpha$  and SREBP1c) that establish and maintain the stable differentiation of the adipocyte. This transcriptional cascade interacts with other transcription factors, co-activators and microRNAs during the decision to proliferate or differentiate and the decision to adopt a brown or beige 'thermogenic' cell fate or a white 'non-thermogenic' cell fate. For example, miR-155 is involved in a feedback loop with C/EBP- $\beta$  that regulates the decision to proliferate or differentiate. During adipogenesis, PPAR- $\alpha$ , PGC1- $\alpha$ , SRC-1, PRDM16, miR-196a and miR-193b-365 drive a thermogenic brown or beige phenotype, whereas RIP140, TIF2, miR-133 drive a white phenotype.

phenotype during adipogenesis. Recent progress using rodent models suggests that determining the brown-, beige- or white-cell fate during adipogenesis requires three levels of control involving transcription factors and co-activators, epigenetics and miRNAs (Fig. 2).

Although PPAR- $\gamma$  is a key driver of the general adipogenic programme, treatment of brown adipocyte precursors with PPAR- $\gamma$  agonists potentiates brown adipogenesis and the same treatment biases murine and human white adipocyte precursors towards a beige-cell fate<sup>20,21,38,39</sup>. Other transcription factors and co-activators are known to drive adipogenesis towards a white or brown fate. For example, the corepressor RIP140 and the co-activator TIF2 help to drive white adipogenesis, whereas the transcription factors PPAR- $\alpha$  and FOXC2 and the PPAR- $\gamma$  co-activators PGC1- $\alpha$ , SRC1 and PRDM16 promote brown adipogenesis<sup>44</sup>. Despite PGC1- $\alpha$ 's importance in regulating the expression of thermogenic genes in brown adipocytes, it also regulates mitochondria biogenesis and oxidative metabolism in many cell types and organs<sup>45</sup>. By contrast, PRDM16 seems to be a more specific driver of thermogenic gene expression. PRDM16 is a co-activator that forms a complex with multiple adipogenic transcription factors, including PPAR- $\gamma$  and PGC1- $\alpha$ <sup>30,46,47</sup>. This complex may drive commitment to a brown adipocyte fate (rather than a muscular cell fate) during brown adipogenesis, but might be more important for beige adipocyte recruitment in WAT<sup>30,46–48</sup>. Whereas adipose-tissue-specific deletion of PRDM16 blocks the induction of thermogenic genes in WAT in response to adrenergic stimulation, paradoxically it has almost no effect on BAT development<sup>48</sup>.

It is also important to understand how specific transcriptional drivers such as PRDM16 and PPAR- $\gamma$  integrate with other mechanisms that regulate cell-fate decisions. For example, epigenetic factors can influence the availability of the promoters targeted by these fate-determining transcription factors. In brown adipocyte precursors, the lysine methyltransferase EHMT1 drives brown adipogenesis by stabilizing PRDM16 and regulating histone methylation at muscle-selective gene promoters<sup>49</sup>. Furthermore, transcription factors can be covalently modified to regulate their activity. This mechanism is exemplified by the deacetylase SIRT1, which targets multiple transcription factors such as PGC1- $\alpha$  and FOXO1 to regulate oxidative metabolism and glucose homeostasis<sup>50–52</sup>.

In addition, SIRT1-mediated deacetylation of PPAR- $\gamma$  favours beige adipocyte recruitment in WAT, and SIRT1-mediated deacetylation of the myogenic transcription factor MyoD might repress the myogenic gene programme during brown adipogenesis<sup>19,31,53,54</sup>.

Finally, miRNAs have recently emerged as regulators of cell differentiation and metabolism in multiple tissues, including BAT and WAT<sup>32,55</sup>. For example, in brown adipocyte precursors miR-155 is involved in a feedback loop with C/EBP- $\beta$  that regulates the decision to proliferate or differentiate<sup>56</sup>. Multiple miRNAs also fine-tune the fate of differentiating WAT and BAT adipocyte precursors. miR-27 is a negative regulator of the brown and beige adipogenic programme in BAT and WAT that is down-regulated by cold exposure, whereas miR-196a expression is increased in WAT precursor adipocytes by cold exposure or  $\beta$ -adrenergic stimulation to induce the recruitment of beige cells<sup>57,58</sup>. In addition, miRNAs coordinate with PRDM16 to regulate the cell-fate decision to become a brown adipocyte rather than a myocyte. In this context, PRDM16 drives brown adipogenesis by inducing the expression of the miRNA cluster miR-193b-365, which targets the brown adipogenic repressor Runx1t1 (ref. 59). Conversely, miR-133 enforces myogenic commitment by repressing PRDM16 expression, and is downregulated in BAT and WAT during brown and beige adipogenesis to allow increased PRDM16 expression<sup>29,60,61</sup>.

### Regulation of adipocyte function in rodents

Therapeutic strategies that increase the number of brown and beige adipocytes by driving adipogenesis towards a thermogenic cell fate might not elicit metabolic benefits unless these thermogenic cells are appropriately activated. Although white, brown and beige adipocytes have divergent roles in terms of energy homeostasis, they are all functionally coordinated by integrating sympathetic nervous and endocrine control.

### The role of the sympathetic nervous system

In times of nutrient excess, energy uptake by white adipocytes is principally regulated by insulin's anabolic role, mediating glucose and lipid uptake, lipogenesis and inhibition of lipolysis<sup>62</sup>. By contrast, the sympathetic nervous system (SNS) regulates responses to energy deficits caused by energy shortages (for example, fasting) or increased energy expenditure (for example, cold exposure)<sup>63</sup>. The physiological response to cold exposure requires an SNS-coordinated crosstalk between white, brown and beige adipocytes with respect to nutrient release, partitioning and utilization (reviewed in refs 9, 64; Fig. 3). For example, increased sympathetic tone to WAT triggers lipolysis so that intracellular lipid stores are released to provide nutrients for other organs, particularly for activated BAT<sup>64</sup>. Increased sympathetic tone to BAT also triggers lipolysis, providing free fatty acids to locally activate UCP1 and to serve as substrates for oxidation<sup>9,64</sup>. Importantly, the substrate uptake machinery is also upregulated, allowing lipids released by WAT to be taken up by BAT<sup>9</sup>. Sustained SNS activation will also result in increased thermogenic capacity due to increased thermogenic gene expression and, in cases of chronic activation, due to proliferation and differentiation of brown adipocyte precursors<sup>9</sup>.

Increased SNS tone to WAT also recruits and activates beige adipocytes, which might contribute to increased total thermogenic capacity<sup>16,17</sup>. The functional importance of beige adipocytes is a burning question that has yet to be compellingly clarified in rodents. This could be because it is difficult to experimentally distinguish the thermogenic contribution of beige adipocytes from that of brown adipocytes, and because the many factors — including cold exposure — that recruit and activate beige adipocytes also recruit and activate brown adipocytes. In support of a physiological role for beige adipocytes in energy balance, murine-strain-dependent differences in resistance to diet-induced obesity have been attributed to the inducibility of beige-adipocyte recruitment in WAT<sup>65</sup>. In addition, increased recruitment of beige adipocytes has been shown to compensate for decreased BAT thermogenesis, and the specific loss of beige adipocytes due to adipose-tissue-specific PRDM16 depletion causes obesity<sup>22,66</sup>.



In any case, because activating thermogenesis could be of therapeutic interest, strategies to optimize SNS-targeted activation might be valuable. However, pharmacological strategies that indiscriminately increase sympathetic signalling result in cardiovascular side effects that are unacceptable when treating obesity<sup>67</sup>. The tissue-selective distribution of specific  $\beta$ -adrenergic receptor ( $\beta$ -AR) isoforms suggested it was feasible to target adipose tissue specifically through the  $\beta$ 3-AR isoform<sup>9,64</sup>. However, whereas in rodents  $\beta$ 3-AR agonists mimic the effects of the cold on BAT and WAT, a human  $\beta$ 3-AR agonist that is both potent and selective remains elusive<sup>68</sup>. This emphasizes the therapeutic relevance of mechanisms that could increase adrenergic responsiveness of thermogenic adipocytes, and of alternative, non-adrenergic pathways that can activate BAT.

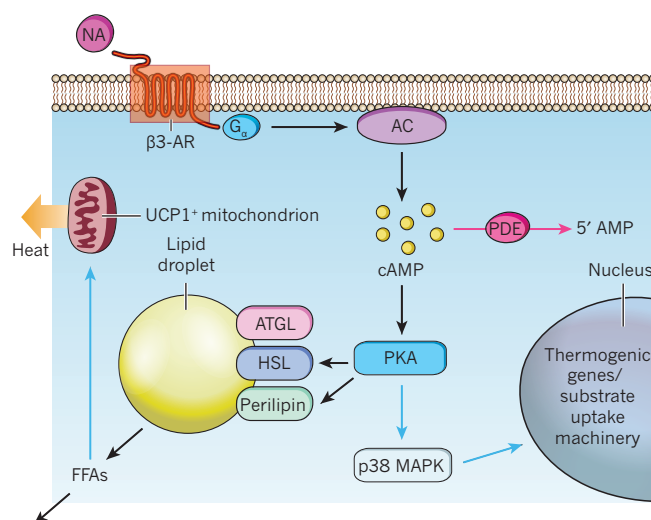
### Alternative pathways that activate thermogenesis

It is possible to directly manipulate the adrenergic signalling pathway to sensitize adipocytes to adrenergic stimulation. For example, the transcription factor FOXC2 sensitizes adipocytes to adrenergic stimulation by modulating the expression and activity of adrenergic signalling molecules<sup>69–71</sup>. In addition, BMP8B is produced by brown adipocytes in response to adrenergic stimulation, and increases the maximal thermogenic response to adrenergic stimulation<sup>72</sup>. BMP8B induces the phosphorylation of downstream adrenergic targets such as hormone-sensitive lipase and p38 MAPK (Fig. 3) by signalling through Smad1, 5 and 8 (ref. 72). This pathway acts in parallel to and synergistically with cAMP and protein kinase A (PKA)-dependent adrenergic signalling<sup>72</sup>. BMP7 can also drive brown adipogenesis and recruit beige adipocytes by activating p38 MAPK<sup>73,74</sup>.

Moreover, several alternative mechanisms that activate brown and beige adipocytes independently of the SNS have been discovered<sup>75</sup>. For example, stimulated brown adipocytes are a significant source of FGF21, a previously recognized liver product that is known to act on the liver and WAT to regulate glucose homeostasis<sup>76</sup>. FGF21 released by brown adipocytes acts locally to further increase BAT thermogenesis, but also enters the circulation and acts systemically to induce beige adipocyte recruitment in WAT by stabilizing PGC-1 $\alpha$  post-translationally<sup>76,77</sup>. In this way FGF21 can coordinate thermogenesis in both BAT and WAT<sup>76,77</sup>. In addition, the adrenergic target PKA shares many intracellular targets with cGMP-dependent protein kinase G (PKG). As a result, cardiac natriuretic peptides (CNP), which act through the cGMP–PKG pathway, can induce lipolysis and the expression of thermogenic genes in brown and white adipocytes, a response which is additive to that of noradrenaline<sup>78,79</sup>. Therefore, like BMP8B signalling, CNP-signalling pathways could be manipulated to boost the responsiveness of brown and beige adipocytes to pre-existing sympathetic tone, or even bypass adrenergic mechanisms altogether to recruit beige adipocytes.

In line with this strategy, the hormone irisin and the cytokine cardiotrophin-1 preferentially elicit a thermogenic response in WAT<sup>18,80,81</sup>. Cardiotrophin-1 is secreted by the heart and binds glycoprotein 130 (gp130), a cytokine receptor that classically activates the JAK–STAT pathway<sup>80,82</sup>. Irisin is produced by white adipocytes and muscle, and acts preferentially on white adipocyte precursors to recruit beige cells by activating the adrenergic target p38 MAPK as well as ERK1/2 (refs 18, 81, 83, 84). The fact that inhibiting ERK signalling reduces the ability of irisin to increase UCP1 expression highlights the importance of this parallel kinase cascade in the thermogenic response<sup>84</sup>.

Finally, the regulation of mitochondrial dynamics is emerging as an important pathway that controls thermogenesis. Adrenergic stimulation of brown adipocytes was recently found to promote uncoupled respiration by inducing mitochondrial fission<sup>85</sup>. FOXC2 also increases thermogenesis, partly by increasing the expression of the mitochondrial transcription factor mtTFA (also known as Tfam), thereby driving mitochondrial biogenesis<sup>71</sup>. Overall, there are several 'alternative' pathways that regulate BAT activity, and that could be explored therapeutically.



**Figure 3 | Regulation of lipolysis and thermogenesis in adipocytes by the sympathetic nervous system.** The sympathetic nervous system triggers intracellular signalling events that lead to lipolysis (black arrows). Noradrenaline (NA) released by sympathetic nerve endings binds to  $\beta$ -adrenoreceptors ( $\beta$ -ARs), which couple to  $G_s$  G-proteins that activate adenylate cyclase (AC), increasing the cellular concentration of cAMP. This activates protein kinase A (PKA), which phosphorylates lipid droplet proteins including hormone-sensitive lipase (HSL) and perilipin, allowing activation of adipose triglyceride lipase (ATGL). The net result is the hydrolysis of triglycerides stored in the lipid droplet to free fatty acids (FFAs), which are released. This response can be attenuated (red arrows) by the activation of phosphodiesterases (PDE), which hydrolyse cAMP. In the case of recruitment and activation of beige adipocytes and the activation of brown adipocytes, this signalling pathway (blue arrows) also leads to increased expression of UCP1 and other thermogenic genes through the p38 MAPK cascade. FFAs released by lipolysis activate existing UCP1 in the mitochondria, and their oxidation produces heat.

### Looking beyond brown and beige adipocytes

The discovery that macrophages within BAT are an alternative source of noradrenaline highlights the fact that BAT consists of more than just brown adipocytes<sup>86</sup>. Given the dependence of optimal BAT activation on innervation and vascularization for gas exchange and heat dissipation, other cell types such as endothelial cells and nerve cells seem to be important contributors to overall BAT function<sup>9</sup>. Therefore, the remodelling of vascular and nervous networks in adipose tissue could enhance the therapeutic benefits of brown and beige adipocyte recruitment and activation. As well as recruiting thermogenic adipocytes, cold exposure also induces angiogenesis and increases the density of sympathetic neurons in adipose tissue<sup>16,87,88</sup>. Furthermore, transgenic rodent models have also demonstrated that thermogenesis and angiogenesis are coupled. Adipose-tissue-specific transgenic expression of FOXC2 in mice results in WAT browning and increased vascularization in WAT due to increased expression of the angiogenic factor angiopoietin 2 (refs 9, 89). In addition, adipose-tissue-specific transgenic expression of VEGF induces angiogenesis, but also increases energy expenditure and induces beige-cell recruitment in WAT<sup>90,91</sup>.

Up to this point, we have predominantly focused on thermogenesis-based strategies that could ameliorate obesity and metabolic lipotoxicity by increasing energy expenditure and disposing of nutrient excess. However, an alternative anti-lipotoxic strategy would be to increase the capacity of WAT for nutrient storage by improving WAT function and enabling WAT expansion. Just as angiogenesis seems to play a key part in BAT expansion and in WAT browning, this process is also important in WAT expansion, in the context of a positive energy balance. In line with this, hypoxia caused by inadequate vascularization of enlarged hypertrophic adipose-tissue depots is one mechanism that drives WAT dysfunction in obesity<sup>92</sup>. Additional mechanisms include altered cytokine

release leading to insulin resistance, immune cell infiltration, inflammation and fibrosis, and are reviewed elsewhere<sup>93</sup>.

Given the deleterious effects of hypoxia on adipocyte function, it is intuitive to suggest that improving the vascularization of WAT could improve WAT function. Experimentally, it seems that either promoting<sup>90,91</sup> or preventing<sup>91,94,95</sup> angiogenesis in adipose tissue can improve the metabolic health of diet- and genetically induced obese mice. One hypothesis to explain this apparent contradiction is that blocking angiogenesis in a state of pre-existing obesity improves WAT function by inducing preferential apoptosis of hypoxic, hypertrophic adipocytes so that only smaller, healthier adipocytes remain, whereas promoting angiogenesis in the early stages of obesity enables healthy WAT expansion<sup>91</sup>.

In line with this, chronic treatment of obese rats with PPAR- $\gamma$  agonists leads to increased weight-gain due to the accumulation of fat mass<sup>96,97</sup>. Paradoxically, this is associated with increased insulin sensitivity<sup>96,97</sup>. WAT expansion can also be achieved by modulating the action of the adipokine adiponectin, by elevating its circulating levels or by adiponectin receptor agonism<sup>98–100</sup>. In diet- and genetically induced obese mice, such treatments further increase body weight and subcutaneous WAT mass, but also drastically improve carbohydrate and lipid metabolism<sup>98–100</sup>. These metabolic benefits are associated with improved WAT function, characterized by decreased adipocyte size (although increased number) and decreased WAT inflammation<sup>98–100</sup>. In an extreme example of facilitating WAT expansion, *ob/ob* mice with adipose-tissue-specific transgenic expression of the mitochondrial protein mitoNEET gain almost twice as much weight as non-transgenic *ob/ob* mice<sup>101</sup>. Nevertheless, these mice have smaller adipocytes in subcutaneous WAT and maintain metabolic parameters comparable with those of wild-type mice<sup>101</sup>. Although reducing adipocyte size is associated with metabolic improvements in obesity, allowing adipocytes to expand freely by knocking out the gene encoding the extracellular matrix component collagen VI also reduces inflammation in WAT and improves whole-body glucose and lipid metabolism<sup>102</sup>. Taken together, these mouse models demonstrate that enabling WAT expansion to buffer nutrient excess can ameliorate the lipotoxic metabolic complications of obesity.

### The relevance of BAT in humans

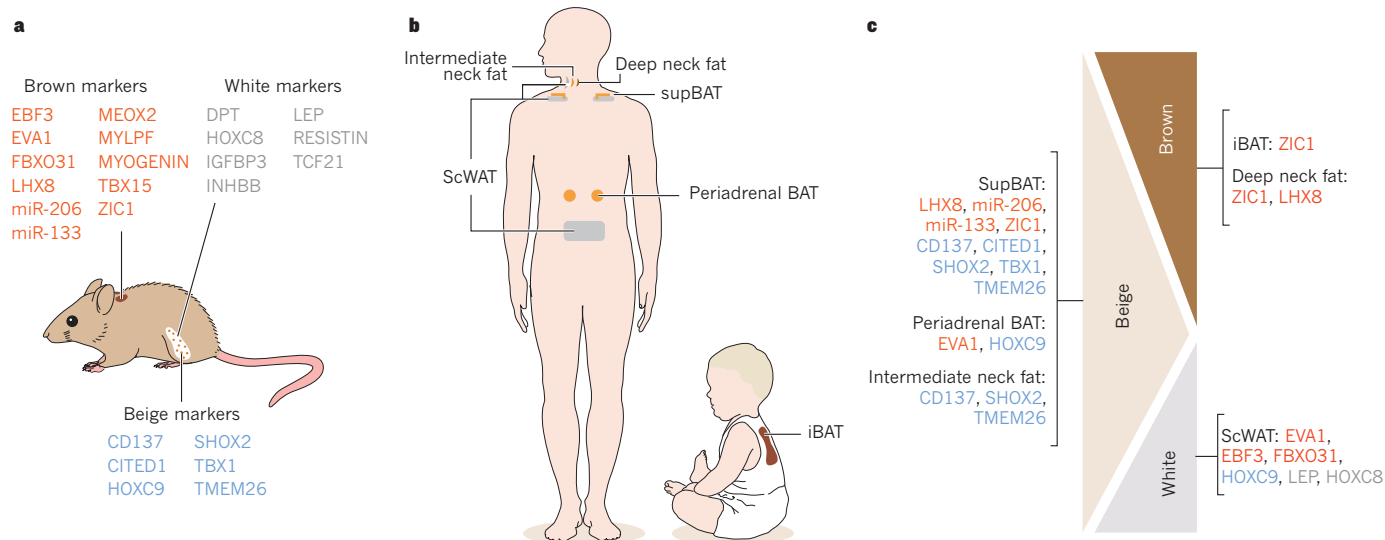
Given that increased fat mass is central to obesity, it is intuitive to target WAT to develop therapeutic strategies for the treatment of metabolic syndrome. Although the use of rodent cellular and *in vivo*

models has greatly informed our current understanding of BAT function and highlighted BAT as a potential anti-obesity target, investigations in human BAT are significantly less advanced.

### Molecular characterization of human BAT

As discussed previously, rodents possess two distinct classes of thermogenic adipocytes. Brown and beige adipocytes have distinct developmental origins and gene expression patterns and are not necessarily optimally stimulated by the same activating molecules. Therefore, determining which types of thermogenic adipocyte are found in human BAT could have profound implications for the successful development of species-specific potent activators. UCP1<sup>+</sup> adipocytes in adult human supraclavicular BAT occur in clusters interspersed among white adipocytes, a pattern that is reminiscent of the beige cells recruited by cold exposure within the subcutaneous WAT depot in rodents<sup>5–8</sup>. On a molecular level, the expression of key thermogenic genes such as *UCP1*, *PGC1A* (also known as *PPARGC1A*) and *PRDM16* in this and other human BAT depots confirms the thermogenic potential of this tissue, but cannot discriminate between a brown-like or beige-like phenotype<sup>5–8,18–21,103–105</sup>.

Thorough molecular comparisons of rodent brown, beige and white adipocytes have identified several candidate discriminatory markers at the messenger RNA level (Supplementary Table 1; Fig. 4)<sup>18–20,31,46,106</sup>. The expression levels of these genes have been measured in multiple human BAT depots in adults and infants<sup>18,20,103–105</sup>. Some rodent brown markers are expressed in human WAT and therefore might not be discriminatory for human thermogenic adipocytes<sup>105</sup>. Nevertheless, overall these results indicate that human BAT depots contain both brown-like and beige-like adipocytes (Supplementary Table 2; Fig. 4)<sup>18,20,103–105</sup>. For example, the infant interscapular BAT (iBAT) depot is dominated by multilocular UCP1<sup>+</sup> adipocytes and separated from surrounding WAT by connective tissue, a structural arrangement like that of the rodent iBAT depot<sup>104</sup>. Molecular characterization of infant iBAT indicates that this depot has higher expression levels of the rodent brown-adipocyte marker *ZIC1* and lower expression levels of the rodent beige-adipocyte marker *TBX1* compared with adult human supraclavicular BAT, suggesting a truly brown identity<sup>104</sup>. In comparison with other adipose-tissue depots, adult human supraclavicular BAT is particularly enriched with markers of rodent beige adipocytes<sup>18,20,103–105</sup> (Fig. 4). However, evidence suggests that adult humans do possess a population of bona fide brown adipocytes in human deep neck fat<sup>105</sup>. In deep neck adipose tissue, *UCP1* expression increases



**Figure 4 | Assessment of rodent brown and beige adipocyte markers in human adipose tissue.** **a**, Well-established rodent brown, beige and white markers. Both white and beige markers are found in subcutaneous white adipose tissue (WAT) depots, but beige markers are upregulated after cold exposure. **b**, Locations of the brown adipose tissue (BAT) and WAT depots

in which the reliability of these rodent markers has been assessed in infants and adults. iBAT, interscapular BAT; supBAT, supraclavicular BAT; scWAT, subcutaneous WAT. **c**, A comparison of the rodent brown, beige and white markers that have been found to be enriched in particular human adipose tissue depots. See Supplementary Tables 1 and 2 for more detailed information.

steeply towards the innermost layers accompanied by elevated expression of the rodent brown-adipocyte markers *ZIC1* and *LHX8* only in the deepest locations<sup>105</sup>. By contrast, layers at intermediate depth with moderate *UCP1* expression are enriched with beige markers<sup>105</sup>.

### Physiological importance of human BAT

Efforts to characterize human BAT on a molecular level have been accompanied by studies investigating the role of BAT in energy balance and metabolism. Estimates of BAT mass and activity from positron emission tomography with <sup>18</sup>F-fluoro-2-deoxy-D-glucose integrated with computed tomography (<sup>18</sup>FDG-PET-CT) suggest that humans have, on average, 50–80 g of BAT<sup>5,107</sup>. This figure divides opinions on whether endogenous BAT activity could be relevant to human energy balance and metabolism. However, <sup>18</sup>FDG-PET-CT does not directly measure either oxidative metabolism or thermogenesis in BAT as it only reports the uptake of the minor oxidative substrate glucose. At room temperature, <sup>18</sup>FDG-PET-CT usually detects BAT in less than 10% of adult humans, but the detection rate increases following mild cold exposure before scanning<sup>6,8</sup>. Therefore, it is possible that <sup>18</sup>FDG-PET-CT studies underestimate the prevalence and volume of BAT in the population, particularly as some subjects do have *UCP1*<sup>+</sup> adipocytes in their supraclavicular BAT despite not accumulating <sup>18</sup>FDG<sup>108</sup>. Furthermore, this technique cannot distinguish clusters of thermogenic adipocytes in WAT<sup>109</sup>.

Despite the limitations of <sup>18</sup>FDG-PET-CT, BAT activity does seem to have a positive relationship with resting metabolic rate (RMR) in healthy adult males, although only at cooler temperatures (22°C compared with 27°C)<sup>6,10</sup>, supporting a role for BAT in whole-body energy expenditure. In addition, multiple studies using this technique report that the likelihood of BAT detection and BAT activity decrease with increasing body mass index, per cent body fat, age and plasma glucose levels<sup>6–8,111–113</sup>. In summary, it would seem that BAT activity is decreased in the metabolically ‘unhealthy’ states of obesity and hyperglycaemia, whereas increased BAT activity is associated with a lean, normoglycaemic, ‘healthy’ metabolic profile. However, in these associations it is unclear which is the cause and which is the consequence.

### Activation and recruitment of human BAT

The short-term, mild-cold-exposure protocol (exposure to temperatures between 15–18°C for up to 6 hours) used by many human BAT studies has been shown to increase the skin temperature over the supraclavicular BAT depot, and to increase glucose uptake, fatty-acid uptake, oxidative metabolism and blood flow in this anatomical area<sup>107,109,114</sup>. These changes are consistent with the activation of human BAT. Furthermore, two recent studies investigated whether chronic cold exposure is able to induce BAT recruitment by measuring BAT activity in response to a single episode of cold exposure before and after cold acclimation<sup>115,116</sup>. Both studies showed that a daily regime of cold exposure — either 2 hours at 17°C daily for 6 weeks or up to 6 hours at 15–16°C daily for 10 days — increased glucose uptake into BAT<sup>115,116</sup>, and one study even reported a 37% increase in estimated BAT volume<sup>115</sup>. Taken together, this is promising evidence that human BAT is an active, highly plastic ‘trainable’ tissue that is responsive to physiological stimuli.

### Cellular models of human adipose tissue

Just as murine cellular models have proved invaluable to the investigation of adipocyte differentiation and functional regulation, human adipocyte cellular models will be key to translating therapeutic strategies to humans. First, *in vitro* culture experiments have succeeded in demonstrating that human BAT is more metabolically active than WAT. The oxygen consumption rates of adipocytes from human supraclavicular BAT and deep neck fat biopsies were found to be higher compared with adipocytes from human subcutaneous WAT biopsies<sup>105,117</sup>. In addition, human cervical fat adipocyte precursors have been differentiated *in vitro* into beige-like cells using a brown adipogenic cocktail<sup>118</sup>. These beige-like cells demonstrated higher basal oxygen consumption rates and increased uncoupling compared with white adipocytes derived from the same precursors using

a white adipogenic cocktail<sup>118</sup>. Furthermore, the oxygen consumption rate of the beige-like cells could be further increased on stimulation with noradrenaline<sup>118</sup>.

Second, *in vitro* models have already been used to demonstrate that alternative activators of rodent thermogenic adipocytes could be relevant to humans. For example, CNPs induce lipolysis and thermogenic gene expression in human white adipocytes<sup>79,119</sup>. In addition, *UCP1* expression and uncoupled respiration can be increased in white adipocytes derived from human subcutaneous WAT and neck fat by treatment with FGF21 or irisin<sup>118,120</sup>. These results suggest that these molecules are physiological thermogenic activators in humans, which is also supported by recent evidence that circulating levels of FGF21 and irisin are increased with cold exposure in humans<sup>120,121</sup>.

Although primary culture and differentiation of human adipocyte precursors is the gold standard for interrogating adipogenic regulatory networks and testing agents that might modulate adipocyte function, these cells are difficult to obtain and have a limited lifetime in culture. Recent advances in stem-cell technology have allowed the field to move towards human embryonic stem cell and induced pluripotent stem (iPS) cell differentiation models, which can provide an unlimited source of human white, brown and beige adipocytes, and their intermediate precursors<sup>122,123</sup>. Such genetically engineerable models are valuable tools for building a comprehensive picture of the regulatory networks that determine brown-, beige- and white-adipocyte fate in humans. Perhaps more importantly, they will enable high-throughput screening of candidate agents for increasing BAT mass and/or activity, and for beige adipocyte recruitment. Furthermore, adipocytes derived from clinical grade iPS cells might even be reintroduced into the body. The use of BAT transplantation as a method to increase BAT mass has already been tested in rodent models, and could be viable given adequate innervation and vascularization of the transplanted material<sup>124,125</sup>. Therefore, cellular models of human adipocytes are valuable tools that might help to bridge the gap between rodents and humans, from drug discovery to transplantation.

### Future perspectives

We propose that the metabolic toxicity that drives the pathology of obesity and metabolic syndrome could be reduced by improving the nutrient storage capacity of WAT and/or by maximizing the energy-dissipating potential of thermogenic brown and beige adipocytes. In our opinion, progress in using BAT to treat metabolic disease requires a better understanding of the molecular nature of the adipocytes in human BAT, which will provide valuable guidance for the translation of strategies to activate brown or beige cells in rodent models into a clinical context. In particular, the development of appropriate human cellular models to enable the screening of potential pharmacological activators is an important first step on the predictably long and winding road to clinical success. ■

Received 18 December 2013; accepted 26 March 2014.

1. Lelliott, C. & Vidal-Puig, A. J. Lipotoxicity, an imbalance between lipogenesis *de novo* and fatty acid oxidation. *Int. J. Obes. Relat. Metab. Disord.* **28**, S22–S28 (2004).
  2. Scherer, P. E. Adipose tissue: from lipid storage compartment to endocrine organ. *Diabetes* **55**, 1537–1545 (2006).
  3. Hany, T. F. et al. Brown adipose tissue: a factor to consider in symmetrical tracer uptake in the neck and upper chest region. *Eur. J. Nucl. Med. Mol. Imaging* **29**, 1393–1398 (2002).
  4. Nedergaard, J. et al. Unexpected evidence for active brown adipose tissue in adult humans. *Am. J. Physiol. Endocrinol. Metab.* **293**, E444–E452 (2007).
  5. Virtanen, K. A. et al. Functional brown adipose tissue in healthy adults. *N. Engl. J. Med.* **360**, 1518–1525 (2009).
  6. van Marken Lichtenbelt, W. D. et al. Cold-activated brown adipose tissue in healthy men. *N. Engl. J. Med.* **360**, 1500–1508 (2009).
  7. Cypess, A. M. et al. Identification and importance of brown adipose tissue in adult humans. *N. Engl. J. Med.* **360**, 1509–1517 (2009).
- Refs 5–7 confirmed the presence of metabolically active BAT in adult humans.**
8. Saito, M. et al. High incidence of metabolically active brown adipose effects of cold exposure and adiposity. *Diabetes* **58**, 1526–1531 (2009).
  9. Cannon, B. & Nedergaard, J. Brown adipose tissue: function and physiological significance. *Physiol. Rev.* **84**, 277–359 (2004).



10. Cannon, B. & Nedergaard, J. Thermogenesis challenges the adipostat hypothesis for body-weight control. *Proc. Nutr. Soc.* **68**, 401–407 (2009).
11. Feldmann, H. M., Golozoubova, V., Cannon, B. & Nedergaard, J. UCP1 ablation induces obesity and abolishes diet-induced thermogenesis in mice exempt from thermal stress by living at thermoneutrality. *Cell Metab.* **9**, 203–209 (2009).
12. Lowell, B. B. *et al.* Development of obesity in transgenic mice after genetic ablation of adipose tissue. *Nature* **366**, 740–742 (1993).
13. Bartelt, A. *et al.* Brown adipose tissue activity controls triglyceride clearance. *Nature Med.* **17**, 200–205 (2011).
14. Arbeeny, C. M., Meyers, D. S., Hillyer, D. E. & Bergquist, K. E. Metabolic alterations associated with the antidiabetic effect of  $\beta$  3-adrenergic receptor agonists in obese mice. *Am. J. Physiol.* **268**, E678–E684 (1995).
15. Rothwell, N. J. & Stock, M. J. Luxuskonsumption, diet-induced thermogenesis and brown fat: the case in favour. *Clin. Sci.* **64**, 19–23 (1983).
16. Loncar, D., Afzelius, B. A. & Cannon, B. Epididymal white adipose tissue after cold stress in rats. I. Nonmitochondrial changes. *J. Ultrastruct. Mol. Struct. Res.* **101**, 109–122 (1988).
17. Young, P., Arch, J. R. & Ashwell, M. Brown adipose tissue in the parametrical fat pad of the mouse. *FEBS Lett.* **167**, 10–14 (1984).
18. Wu, J. *et al.* Beige adipocytes are a distinct type of thermogenic fat cell in mouse and human. *Cell* **150**, 366–376 (2012).
- This major study investigated the brown or beige phenotype of the adipocytes found in human supraclavicular BAT, and included a comprehensive gene expression analysis comparing rodent brown, beige and white adipocytes.**
19. Waldén, T. B. *et al.* Recruited vs. nonrecruited molecular signatures of brown, "brite," and white adipose tissues. *Am. J. Physiol. Endocrinol. Metab.* **302**, E19–E31 (2012).
20. Sharp, L. Z. *et al.* Human BAT possesses molecular signatures that resemble beige/brite cells. *PLoS ONE* **7**, e49452 (2012).
21. Petrovic, N. *et al.* Chronic peroxisome proliferator-activated receptor  $\gamma$  (PPAR- $\gamma$ ) activation of epididymally derived white adipocyte cultures reveals a population of thermogenically competent, UCP1-containing adipocytes molecularly distinct from classic brown adipocyte. *J. Biol. Chem.* **285**, 7153–7164 (2010).
22. Shabalina, I. G. *et al.* UCP1 in brite/beige adipose tissue mitochondria is functionally thermogenic. *Cell Rep.* **5**, 1196–1203 (2013).
23. Virtue, S. & Vidal-Puig, A. Adipose tissue expandability, lipotoxicity and the metabolic syndrome—an allostatic perspective. *Biochim. Biophys. Acta* **1801**, 338–349 (2010).
24. Rodeheffer, M. S., Birsoy, K. & Friedman, J. M. Identification of white adipocyte progenitor cells *in vivo*. *Cell* **135**, 240–249 (2008).
25. Tang, W. *et al.* White fat progenitor cells reside in the adipose vasculature. *Science* **322**, 583–586 (2008).
26. Carobbio, S., Rosen, B. & Vidal-Puig, A. Adipogenesis: new insights into brown adipose tissue differentiation. *J. Mol. Endocrinol.* **51**, T75–T85 (2013).
27. Seale, P. *et al.* Pax7 is required for the specification of myogenic satellite cells. *Cell* **102**, 777–786 (2000).
28. Lepper, C. & Fan, C. Inducible lineage tracing of Pax7-descendant cells reveals embryonic origin of adult satellite cells. *Genesis* **48**, 424–436 (2010).
29. Yin, H. *et al.* MicroRNA-133 controls brown adipose determination in skeletal muscle satellite cells by targeting Prdm16. *Cell Metab.* **17**, 210–224 (2013).
30. Seale, P. *et al.* PRDM16 controls a brown fat/skeletal muscle switch. *Nature* **454**, 961–967 (2008).
- This study included a fate-mapping approach to identify the lineage origin of brown and white adipocytes, and identified PRDM16 as a 'cell-fate switch' that specifies a brown fat or muscle cell fate.**
31. Timmons, J. A. *et al.* Myogenic gene expression signature establishes that brown and white adipocytes originate from distinct cell lineages. *Proc. Natl Acad. Sci. USA* **104**, 4401–4406 (2007).
32. Walden, T. B., Timmons, J. A., Keller, P., Nedergaard, J. & Cannon, B. Distinct expression of muscle-specific microRNAs (myomirs) in brown adipocytes. *J. Cell. Physiol.* **218**, 444–449 (2009).
33. Sanchez-Gurmaches, J. *et al.* PTEN loss in the Myf5 lineage redistributes body fat and reveals subsets of white adipocytes that arise from Myf5 precursors. *Cell Metab.* **16**, 348–362 (2012).
34. Shan, T. *et al.* Distinct populations of adipogenic and myogenic Myf5-lineage progenitors in white adipose tissues. *J. Lipid Res.* **54**, 2214–2224 (2013).
35. Tran, K. V. *et al.* The vascular endothelium of the adipose tissue give rise to both white and brown fat cells. *Cell Metab.* **15**, 222–229 (2012).
36. Gupta, R. K. *et al.* Zfp423 expression identifies committed preadipocytes and localizes to adipose endothelial perivascular cells. *Cell Metab.* **15**, 230–239 (2012).
37. Berry, R. & Rodeheffer, M. S. Characterization of the adipocyte cellular lineage *in vivo*. *Nature Cell Biol.* **15**, 302–308 (2013).
38. Digby, J. E. *et al.* Thiazolidinedione exposure increases the expression of uncoupling protein 1 in cultured human preadipocytes. *Diabetes* **47**, 138–141 (1998).
39. Petrovic, N., Shabalina, I. G., Timmons, J. A., Cannon, B. & Nedergaard, J. Thermogenically competent nonadrenergic recruitment in brown preadipocytes by a PPAR $\gamma$  agonist. *Am. J. Physiol. Endocrinol. Metab.* **295**, E287–E296 (2008).
40. Rosenwald, M., Perdikari, A., Rüllicke, T. & Wolfrum, C. Bi-directional interconversion of brite and white adipocytes. *Nature Cell Biol.* **15**, 659–667 (2013).
41. Wang, Q. A., Tao, C., Gupta, R. K. & Scherer, P. E. Tracking adipogenesis during white adipose tissue development, expansion and regeneration. *Nature Med.* **19**, 1338–1344 (2013).
- Refs 40 and 41 both used genetically engineered mouse models to identify the adult origins of beige cells recruited by cold exposure.**
42. Lee, Y.-H., Petkova, A. P., Mottillo, E. P. & Grannem, J. G. *In vivo* identification of bipotential adipocyte progenitors recruited by  $\beta$ 3-adrenoceptor activation and high-fat feeding. *Cell Metab.* **15**, 480–491 (2012).
43. Liu, W. *et al.* A heterogeneous lineage origin underlies the phenotypic and molecular differences of white and beige adipocytes. *J. Cell Sci.* **126**, 3527–3532 (2013).
44. Hansen, J. B. & Kristiansen, K. Regulatory circuits controlling white versus brown adipocyte differentiation. *Biochem. J.* **398**, 153–168 (2006).
45. Puigserver, P. Tissue-specific regulation of metabolic pathways through the transcriptional coactivator PGC1- $\alpha$ . *Int. J. Obes.* **29**, S5–S9 (2005).
46. Seale, P. *et al.* Transcriptional control of brown fat determination by PRDM16. *Cell Metab.* **6**, 38–54 (2007).
47. Seale, P. *et al.* Prdm16 determines the thermogenic program of subcutaneous white adipose tissue in mice. *J. Clin. Invest.* **121**, 96–105 (2011).
48. Cohen, P. *et al.* Ablation of PRDM16 and beige adipose causes metabolic dysfunction and a subcutaneous to visceral fat switch. *Cell* **156**, 304–316 (2014).
49. Ohno, H., Shinoda, K., Ohyama, K., Sharp, L. Z. & Kajimura, S. EHMT1 controls brown adipose cell fate and thermogenesis through the PRDM16 complex. *Nature* **504**, 163–167 (2013).
50. Frescas, D., Valenti, L. & Accili, D. Nuclear trapping of the forkhead transcription factor FoxO1 via Sirt-dependent deacetylation promotes expression of glucogenic genes. *J. Biol. Chem.* **280**, 20589–20595 (2005).
51. Rodgers, J. T. *et al.* Nutrient control of glucose homeostasis through a complex of PGC-1 $\alpha$  and SIRT1. *Nature* **434**, 113–118 (2005).
52. Lagogue, M. *et al.* Resveratrol improves mitochondrial function and protects against metabolic disease by activating SIRT1 and PGC-1 $\alpha$ . *Cell* **127**, 1109–1122 (2006).
53. Fulco, M. *et al.* Sir2 regulates skeletal muscle differentiation as a potential sensor of the redox state. *Mol. Cell* **12**, 51–62 (2003).
54. Qiang, L. *et al.* Brown remodeling of white adipose tissue by SirT1-dependent deacetylation of Ppar $\gamma$ . *Cell* **150**, 620–632 (2012).
55. Trajkovski, M. & Lodish, H. MicroRNA networks regulate development of brown adipocytes. *Trends Endocrinol. Metab.* **24**, 442–450 (2013).
56. Chen, Y. *et al.* miR-155 regulates differentiation of brown and beige adipocytes via a bistable circuit. *Nature Commun.* **4**, 1769 (2013).
57. Sun, L. & Trajkovski, M. MiR-27 orchestrates the transcriptional regulation of brown adipogenesis. *Metabolism* **63**, 272–282 (2014).
58. Mori, M., Nakagami, H., Rodriguez-Araujo, G., Nimura, K. & Kaneda, Y. Essential role for miR-196a in brown adipogenesis of white fat progenitor cells. *PLoS Biol.* **10**, e1001314 (2012).
59. Sun, L. *et al.* MiR-193b-365, a brown fat enriched microRNA cluster, is essential for brown fat differentiation. *Nature Cell Biol.* **13**, 958–965 (2011).
60. Trajkovski, M., Ahmed, K., Esau, C. C. & Stoffel, M. MyomiR-133 regulates brown fat differentiation through Prdm16. *Nature Cell Biol.* **14**, 1330–1335 (2012).
61. Liu, W. *et al.* miR-133a regulates adipocyte browning *in vivo*. *PLoS Genet.* **9**, e1003626 (2013).
62. Wilcox, G. Insulin and insulin resistance. *Clin. Biochem. Rev.* **26**, 19–39 (2005).
63. Geering, J. J. *et al.* Sympathetic nervous system control of triglyceride metabolism: novel concepts derived from recent studies. *J. Lipid Res.* **55**, 180–189 (2014).
64. Collins, S. & Surwit, R. S. The  $\beta$ -adrenergic receptors and the control of adipose tissue metabolism and thermogenesis. *Recent Prog. Horm. Res.* **56**, 309–328 (2001).
65. Schulz, T. J. & Tseng, Y.-H. Brown adipose tissue: development, metabolism and beyond. *Biochem. J.* **453**, 167–178 (2013).
66. Schulz, T. J. *et al.* Brown-fat paucity due to impaired BMP signalling induces compensatory browning of white fat. *Nature* **495**, 379–383 (2013).
67. Torp-Pedersen, C. *et al.* Cardiovascular responses to weight management and sibutramine in high-risk subjects: an analysis from the SCOUT trial. *Eur. Heart J.* **28**, 2915–2923 (2007).
68. Arch, J. R. S.  $\beta$ 3-adrenoceptor agonists: potential, pitfalls and progress. *Eur. J. Pharmacol.* **440**, 99–107 (2002).
69. Cederberg, A. *et al.* FOXC2 is a winged helix gene that counteracts obesity, hypertriglyceridemia, and diet-induced insulin resistance. *Cell* **106**, 563–573 (2001).
70. Grønning, L. M. *et al.* Reduced PDE4 expression and activity contributes to enhanced catecholamine-induced cAMP accumulation in adipocytes from FOXC2 transgenic mice. *FEBS Lett.* **580**, 4126–4130 (2006).
71. Lidell, M. E. *et al.* The adipocyte-expressed forkhead transcription factor Foxc2 regulates metabolism through altered mitochondrial function. *Diabetes* **60**, 427–435 (2011).
72. Whittle, A. J. *et al.* BMP8B increases brown adipose tissue thermogenesis through both central and peripheral actions. *Cell* **149**, 871–885 (2012).
73. Tseng, Y.-H. *et al.* New role of bone morphogenetic protein 7 in brown adipogenesis and energy expenditure. *Nature* **454**, 1000–1004 (2008).
74. Boon, M. R. *et al.* BMP7 activates brown adipose tissue and reduces diet-induced obesity only at subthermoneutrality. *PLoS ONE* **8**, e74083 (2013).
75. Villarroya, F. & Vidal-Puig, A. Beyond the sympathetic tone: the new brown fat activators. *Cell Metab.* **17**, 638–643 (2013).
76. Hondares, E. *et al.* Thermogenic activation induces FGF21 expression and release in brown adipose tissue. *J. Biol. Chem.* **286**, 12983–12990 (2011).
77. Fisher, F. M. *et al.* FGF21 regulates PGC-1 $\alpha$  and browning of white adipose tissues in adaptive thermogenesis. *Genes Dev.* **26**, 271–281 (2012).

78. Lafontan, M. *et al.* Control of lipolysis by natriuretic peptides and cyclic GMP. *Trends Endocrinol. Metab.* **19**, 130–137 (2008).
79. Bordicchia, M. *et al.* Cardiac natriuretic peptides act via p38 MAPK to induce the brown fat thermogenic program in mouse and human adipocytes. *J. Clin. Invest.* **122**, 1022–1036 (2012).
80. Moreno-Aliaga, M. J. *et al.* Cardiotrophin-1 is a key regulator of glucose and lipid metabolism. *Cell Metab.* **14**, 242–253 (2011).
81. Boström, P. *et al.* A PGC1- $\alpha$ -dependent myokine that drives brown-fat-like development of white fat and thermogenesis. *Nature* **481**, 463–468 (2012).
82. Jones, S. A., Scheller, J. & Rose-John, S. Therapeutic strategies for the clinical blockade of IL-6 / gp130 signaling. *J. Clin. Invest.* **121**, 3375–3383 (2011).
83. Roca-Rivada, A. *et al.* FNDC5/irisin is not only a myokine but also an adipokine. *PLoS ONE* **8**, e60563 (2013).
84. Zhang, Y. *et al.* Irisin stimulates browning of white adipocytes through mitogen-activated protein kinase p38 MAP kinase and ERK MAP kinase signaling. *Diabetes* **63**, 514–525 (2014).
85. Wikström, J. D. *et al.* Hormone-induced mitochondrial fission is utilized by brown adipocytes as an amplification pathway for energy expenditure. *EMBO J.* **33**, 418–436 (2014).
86. Nguyen, K. D. *et al.* Alternatively activated macrophages produce catecholamines to sustain adaptive thermogenesis. *Nature* **480**, 104–108 (2011).
87. Xue, Y. *et al.* Hypoxia-independent angiogenesis in adipose tissues during cold acclimation. *Cell Metab.* **9**, 99–109 (2009).
88. Vitali, A. *et al.* The adipose organ of obesity-prone C57BL/6J mice is composed of mixed white and brown adipocytes. *J. Lipid Res.* **53**, 619–629 (2012).
89. Xue, Y. *et al.* FOXO2 controls Ang-2 expression and modulates angiogenesis, vascular patterning, remodeling, and functions in adipose tissue. *Proc. Natl Acad. Sci. USA* **105**, 10167–10172 (2008).
90. Elias, I. *et al.* Adipose tissue overexpression of vascular endothelial growth factor protects against diet-induced obesity and insulin resistance. *Diabetes* **61**, 1801–1813 (2012).
91. Sun, K. *et al.* Dichotomous effects of VEGF-A on adipose tissue dysfunction. *Proc. Natl Acad. Sci. USA* **109**, 5874–5879 (2012).
- The authors of this study dissected the contribution of angiogenesis to WAT function and dysfunction during the development of obesity compared with a state of existing obesity.**
92. Trayhurn, P., Wang, B. & Wood, I. S. Hypoxia and the endocrine and signalling role of white adipose tissue. *Arch. Physiol. Biochem.* **114**, 267–276 (2008).
93. Ouchi, N., Parker, J. L., Lugus, J. J. & Walsh, K. Adipokines in inflammation and metabolic disease. *Nature Rev. Immunol.* **11**, 85–97 (2011).
94. Kolonin, M. G., Saha, P. K., Chan, L., Pasqualini, R. & Arap, W. Reversal of obesity by targeted ablation of adipose tissue. *Nature Med.* **10**, 625–632 (2004).
95. Rupnick, M. A. *et al.* Adipose tissue mass can be regulated through the vasculature. *Proc. Natl Acad. Sci. USA* **99**, 10730–10735 (2002).
96. Okuno, A. *et al.* Troglitazone increases the number of small adipocytes without the change of white adipose tissue mass in obese Zucker rats. *J. Clin. Invest.* **101**, 1354–1361 (1998).
97. De Souza, C. J. *et al.* Effects of pioglitazone on adipose tissue remodeling within the setting of obesity and insulin resistance. *Diabetes* **50**, 1863–1871 (2001).
98. Kim, J. Y. *et al.* Obesity-associated improvements in metabolic profile through expansion of adipose tissue. *J. Clin. Invest.* **117**, 2621–2637 (2007).
99. Yamauchi, T. *et al.* The fat-derived hormone adiponectin reverses insulin resistance associated with both lipoatrophy and obesity. *Nature Med.* **7**, 941–946 (2001).
100. De Souza, C. J. *et al.* Effects of pioglitazone on adipose tissue remodeling within the setting of obesity and insulin resistance. *Diabetes* **50**, 1863–1871 (2001).
101. Kusminski, C. M. *et al.* MitoNEET-driven alterations in adipocyte mitochondrial activity reveal a crucial adaptive process that preserves insulin sensitivity in obesity. *Nature Methods* **18**, 1539–1549 (2012).
102. Khan, T. *et al.* Metabolic dysregulation and adipose tissue fibrosis: role of collagen VI. *Mol. Cell. Biol.* **29**, 1575–1591 (2009).
103. Jespersen, N. Z. *et al.* A classical brown adipose tissue mRNA signature partly overlaps with white in the supraclavicular region of adult humans. *Cell Metab.* **17**, 798–805 (2013).
104. Lidell, M. E. *et al.* Evidence for two types of brown adipose tissue in humans. *Nature Med.* **19**, 631–634 (2013).
- Infant interscapular BAT was found to express markers of canonical rodent brown adipocytes, indicating the presence of bona fide brown adipose tissue in humans.**
105. Cypess, A. M. *et al.* Anatomical localization, gene expression profiling and functional characterization of adult human neck brown fat. *Nature Med.* **19**, 635–639 (2013).
- In this gene expression analysis, adult human neck fat showed enrichment of rodent brown-adipocyte markers and rodent beige-adipocyte markers in distinct locations, suggesting the co-existence of brown and beige-like adipocytes in adult humans.**
106. Cinti, S. *et al.* Immunohistochemical localization of leptin and uncoupling protein in white and brown adipose tissue. *Endocrinology* **138**, 797–804 (1997).
107. Ouellet, V. *et al.* Brown adipose tissue oxidative metabolism contributes to energy expenditure during acute cold exposure in humans. *J. Clin. Invest.* **122**, 545–552 (2012).
108. Lee, P. *et al.* High prevalence of brown adipose tissue in adult humans. *J. Clin. Endocrinol. Metab.* **96**, 2450–2455 (2011).
109. Muzik, O. *et al.*  $^{18}\text{O}$  PET measurement of blood flow and oxygen consumption in cold-activated human brown fat. *J. Nucl. Med.* **54**, 523–531 (2013).
110. Yoneshiro, T. *et al.* Brown adipose tissue, whole-body energy expenditure, and thermogenesis in healthy adult men. *Obesity (Silver Spring)* **19**, 13–16 (2011).
111. Pfannenberger, C. *et al.* Impact of age on the relationships of brown adipose tissue with sex and adiposity in humans. *Diabetes* **59**, 1789–1793 (2010).
112. Ouellet, V. *et al.* Outdoor temperature, age, sex, body mass index, and diabetic status determine the prevalence, mass, and glucose-uptake of  $^{18}\text{F}$ -FDG-detected BAT in humans. *J. Clin. Endocrinol. Metab.* **96**, 192–199 (2011).
113. Yoneshiro, T. *et al.* Age-related decrease in cold-activated brown adipose tissue and accumulation of body fat in healthy humans. *Obesity (Silver Spring)* **19**, 1755–1760 (2011).
114. Symonds, M. E. *et al.* Thermal imaging to assess age-related changes of skin temperature within the supraclavicular region co-locating with brown adipose tissue in healthy children. *J. Pediatr.* **161**, 892–898 (2012).
115. van der Lans, A. A. J. J. *et al.* Cold acclimation recruits human brown fat and increases nonshivering thermogenesis. *J. Clin. Invest.* **123**, 3395–3403 (2013).
116. Yoneshiro, T. *et al.* Recruited brown adipose tissue as an antiobesity agent in humans. *J. Clin. Invest.* **123**, 3404–3408 (2013).
117. Vijgen, G. H. E. J. *et al.* Increased oxygen consumption in human adipose tissue from the “brown adipose tissue” region. *J. Clin. Endocrinol. Metab.* **98**, E1230–E1234 (2013).
118. Lee, P., Werner, C. D., Kebebew, E. & Celi, F. S. Functional thermogenic beige adipogenesis is inducible in human neck fat. *Int. J. Obes.* **38**, 170–176 (2014).
119. Sengenès, C., Berlan, M., De Gieseinski, I., Lafontan, M. & Galitzky, J. Natriuretic peptides: a new lipolytic pathway in human adipocytes. *FASEB J.* **14**, 1345–1351 (2000).
120. Lee, P. *et al.* Irisin and FGF21 are cold-induced endocrine activators of brown fat function in humans. *Cell Metab.* **19**, 302–309 (2014).
121. Lee, P. *et al.* Mild cold exposure modulates fibroblast growth factor 21 (FGF21) diurnal rhythm in humans: relationship between FGF21 levels, lipolysis, and cold-induced thermogenesis. *J. Clin. Endocrinol. Metab.* **98**, E98–E102 (2013).
122. Ahfeldt, T. *et al.* Programming human pluripotent stem cells into white and brown adipocytes. *Nature Cell Biol.* **14**, 209–219 (2012).
123. Nishio, M. *et al.* Production of functional classical brown adipocytes from human pluripotent stem cells using specific hemopoietin cocktail without gene transfer. *Cell Metab.* **16**, 394–406 (2012).
- Refs 122 and 123 are pioneering studies in which brown adipocytes were generated from human embryonic stem cells and human induced pluripotent stem cells.**
124. Gunawardana, S. C. & Piston, D. W. Reversal of type 1 diabetes in mice by brown adipose tissue transplant. *Diabetes* **61**, 674–682 (2012).
125. Stanford, K. I. *et al.* Brown adipose tissue regulates glucose homeostasis and insulin sensitivity. *J. Clin. Invest.* **123**, 215–223 (2013).

**Supplementary Information** is linked to the online version of this paper at [www.nature.com/tlsy4e](http://www.nature.com/tlsy4e).

**Acknowledgments** We apologize to those whose work was not directly cited because of space constraints. The authors are supported by FP7 BetaBAT, BBSRC, BHF and MRC programme grants, Wellcome Trust, and Cambridge Overseas Trust.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at [go.nature.com/tlsy4e](http://go.nature.com/tlsy4e). Correspondence should be addressed to V.P. ([vm272@medschl.cam.ac.uk](mailto:vm272@medschl.cam.ac.uk)) and A.V.-P. ([ajv22@medschl.cam.ac.uk](mailto:ajv22@medschl.cam.ac.uk)).

# The role of hepatic lipids in hepatic insulin resistance and type 2 diabetes

Rachel J. Perry<sup>1</sup>, Varman T. Samuel<sup>1,2</sup>, Kitt F. Petersen<sup>1,3</sup> & Gerald I. Shulman<sup>1,3,4,5</sup>

**Non-alcoholic fatty liver disease and its downstream sequelae, hepatic insulin resistance and type 2 diabetes, are rapidly growing epidemics, which lead to increased morbidity and mortality rates, and soaring health-care costs. Developing interventions requires a comprehensive understanding of the mechanisms by which excess hepatic lipid develops and causes hepatic insulin resistance and type 2 diabetes. Proposed mechanisms implicate various lipid species, inflammatory signalling and other cellular modifications. Studies in mice and humans have elucidated a key role for hepatic diacylglycerol activation of protein kinase C $\epsilon$  in triggering hepatic insulin resistance. Therapeutic approaches based on this mechanism could alleviate the related epidemics of non-alcoholic fatty liver disease and type 2 diabetes.**

Modern global health care faces challenges that are drastically different from past generations, largely owing to the increasing worldwide prevalence of obesity. This is exemplified by a change in focus to centre on obesity-related liver disease. Although viral hepatitis continues to be an important health concern, non-alcoholic fatty liver disease (NAFLD) is the now most common liver disorder in the Western world, where the rates of adult and paediatric obesity have soared to an estimated 20–30% of the US population<sup>1,2</sup>. In east and south Asian communities, NAFLD is also on the rise, with estimates that its prevalence reaches as high as 60% in urban areas<sup>3,4</sup>. Startlingly, NAFLD has been found to be highly prevalent among young lean south Asian Indians<sup>5,6</sup>.

A strong association between NAFLD and type 2 diabetes has been demonstrated: more than 90% of obese patients with type 2 diabetes have NAFLD<sup>7</sup>. Insulin resistance is common in both conditions<sup>5</sup>. Patients with NAFLD almost universally have hepatic insulin resistance, which increases the risk of impaired fasting glucose and type 2 diabetes<sup>5,8–11</sup>. In addition, a subset of patients with NAFLD will develop non-alcoholic steatohepatitis (NASH) with histological changes such as steatosis, lobular inflammation and/or hepatocellular ballooning<sup>12</sup>. Around 20% of patients with NASH will progress to liver cirrhosis and liver failure<sup>13,14</sup>. NASH-associated cirrhosis is now the third most common indication for liver transplantation in the United States<sup>15</sup>. Health policies that can prevent NAFLD and new treatments that can reverse the disease will offer tremendous benefits, in terms of both lives saved and health-care costs.

Thus, in this Perspective we will discuss the link between hepatic lipid accumulation and hepatic insulin resistance and focus on the role of diacylglycerol, a lipid metabolite that activates novel protein kinase C isoforms (PKCs) and thereby impairs insulin signalling, in the pathogenesis of lipid-induced hepatic insulin resistance. Although several other mechanisms have been proposed to explain this association, these alternatives have been reviewed elsewhere<sup>16</sup>. As we will discuss here, diacylglycerol-induced novel PKC activation has emerged as a common mechanism to explain the development of insulin resistance in liver and skeletal muscle in a variety of experimental and clinical models.

## Molecular mechanism of lipid-induced insulin resistance

Insulin action requires a coordinated, intricate relay of intracellular signals, involving mostly phosphorylation and dephosphorylation events.

In the canonical view of hepatic insulin signalling, insulin binds and activates the insulin receptor tyrosine kinase (IRTK), which in turn promotes tyrosine kinase phosphorylation of insulin receptor substrates (IRS), most importantly IRS2 in the liver (Fig. 1)<sup>17</sup>. Phosphorylation of IRS2 generates binding sites for Src homology 2 domain proteins, including phosphatidylinositol-3-OH kinase (PI(3)K)<sup>18</sup>. The binding of PI(3)K to IRS2 recruits phosphatidylinositol-3,4,5-trisphosphate (PtdIns(3,4,5)P<sub>3</sub>), which in turn recruits Akt<sup>19</sup>. Under insulin-stimulated conditions, 3-phosphoinositide-dependent kinase-1 phosphorylates and activates Akt, which is thought to suppress hepatic glucose production through two key mechanisms: first, decreased expression of gluconeogenic enzymes by phosphorylation and nuclear exclusion of the forkhead box protein FOXO1 and its pro-gluconeogenic targets, and second, activation of glycogen synthase by phosphorylation and inactivation of glycogen synthase kinase-3 $\beta$ . Although this relatively linear construct is useful for interrogating insulin signalling in experimental models, it fails to capture the interwoven mechanisms that have evolved to regulate hepatic glucose and lipid metabolism. For example, although acute insulin signalling following a meal can decrease messenger RNA expression of gluconeogenic enzymes, it probably does not acutely alter the protein levels of these enzymes. Gluconeogenic enzymes are also conventionally thought to be subject to allosteric activation: acetyl coenzyme A (acetyl-CoA) activates pyruvate carboxylase<sup>20,21</sup>, and fructose-2,6-bisphosphate inhibits fructose-1,6-bisphosphatase<sup>22</sup>. And, although insulin might activate glycogen synthesis, glucose is necessary to inhibit glycogen phosphorylase and effectively promote net hepatic glycogen synthesis<sup>23,24</sup>.

The development of NAFLD is strongly associated with hepatic insulin resistance. This relationship is most apparent when NAFLD is induced in rats after just 3 days of being fed a high-fat diet. The ability of insulin to suppress hepatic glucose production is diminished in this model even without changes in body weight, adiposity or muscle insulin resistance<sup>25</sup>. Hepatic insulin resistance in this model was associated with increased hepatic diacylglycerol content and increased translocation of the primary novel PKC isoform in liver, protein kinase-C $\epsilon$  (PKC $\epsilon$ )<sup>26,27</sup>, to the plasma membrane at which it was found to bind and inhibit the activity of the intracellular kinase domain of the insulin receptor. This was associated with reduced insulin-stimulated phosphorylation of IRS2 and IRS2-associated PI(3)K activity and phosphorylation of Akt2. Consequently,

<sup>1</sup>Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut 06520, USA. <sup>2</sup>VA Connecticut Healthcare System West Haven, Connecticut 06516, USA. <sup>3</sup>Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen DK-2200, Denmark. <sup>4</sup>Department of Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, Connecticut 06520, USA. <sup>5</sup>Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut 06535-8012, USA.



the ability of insulin to activate glycogen synthesis and inhibit gluconeogenesis was impaired (Fig. 1)<sup>25</sup>. The crucial role of PKC $\epsilon$  in mediating lipid-induced hepatic insulin resistance has been convincingly demonstrated by knocking down expression of PKC $\epsilon$  in the liver of rodents. Rats treated with an antisense oligonucleotide (ASO) to decrease hepatic expression of PKC $\epsilon$  were protected from lipid-induced hepatic insulin resistance, although hepatic diacylglycerol and triglyceride content were unchanged<sup>26</sup>. Furthermore, these animals were found to have preserved IRTK activity with intact signalling through downstream proteins. Similarly, *Prkce* whole-body knockout mice are also protected from lipid-induced insulin resistance<sup>28</sup>.

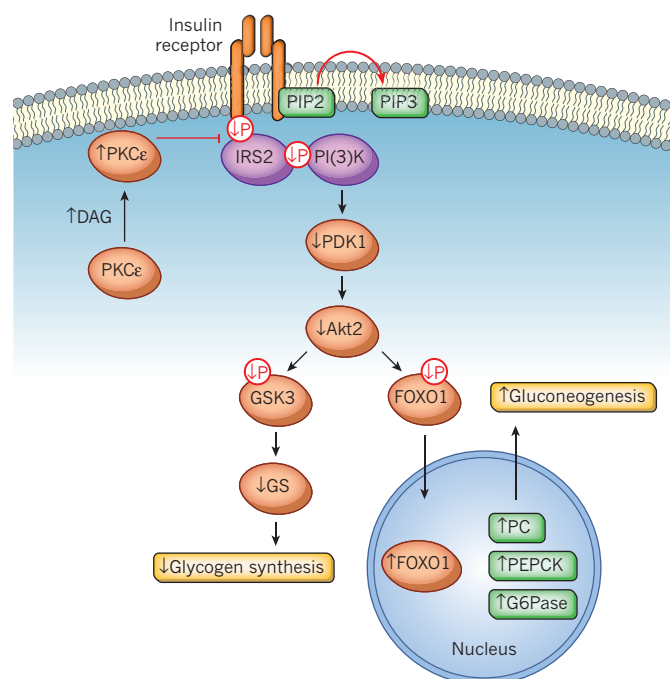
This model for lipid-induced hepatic insulin resistance has been translated to humans. Potential mechanisms for hepatic insulin resistance were assessed in a group of patients undergoing bariatric surgery. Although the participants were all obese, there was a significant variation in insulin resistance. Notably, individuals with very similar body mass index could manifest markedly different degrees of insulin resistance. By contrast, hepatic diacylglycerol content and PKC $\epsilon$  activation were the strongest predictors of hepatic insulin resistance in liver biopsies obtained from these individuals<sup>29</sup>. There was no association between insulin sensitivity and other factors implicated in causing hepatic insulin resistance, including ceramide content, endoplasmic reticulum (ER) stress markers or inflammatory cytokine concentrations. These results were replicated in another study showing that hepatic diacylglycerol content was the best predictor of hepatic insulin resistance in obese humans, whereas there was no association with hepatic ceramide content or markers of inflammation<sup>30</sup>. Indeed, hepatic inflammation has been suggested to be a consequence, not a cause, of insulin resistance<sup>31</sup>. Thus, although excess calorie intake certainly leads to obesity, only those who develop hepatic steatosis will develop insulin resistance. These data argue that the key step in the pathogenesis of hepatic insulin resistance is the accumulation of hepatic diacylglycerol leading to activation of PKC $\epsilon$ . Experimental models of altered hepatic lipid content allow us to further test this mechanism and could also inform the development of potential therapeutic interventions.

The model outlined focuses on the inability of insulin to alter hepatic glucose metabolism. Despite this, the ability of insulin to activate lipogenesis seems to be intact in most models of NAFLD. Much has been written about the paradox of selective insulin resistance, with investigators proposing the existence of branch points in the insulin signalling pathway<sup>32,33</sup>, but space limitations preclude discussion of selective insulin resistance in this Perspective. Instead, we will discuss the mechanisms that govern hepatic lipid accumulation and its relationship to insulin's ability to control hepatic glucose metabolism.

### Regulation of fat delivery to liver

Hepatic lipid content is regulated by the balance between hepatic lipid uptake, synthesis, oxidation and export (Fig. 2). Hepatic lipid uptake is a function of substrate delivery and transport into the hepatocyte, and several genetic models exemplify this aspect of hepatic lipid metabolism. Transgenic mice with liver-specific overexpression of lipoprotein lipase (LpL) develop liver-specific lipid accumulation and liver-specific insulin resistance, whereas transgenic mice with muscle-specific overexpression of LpL develop muscle-specific lipid accumulation and muscle-specific insulin resistance<sup>34</sup>. In these models, tissue-specific insulin resistance followed ectopic lipid accumulation. In a similar example, mice that lack the primary fatty-acid transporter in hepatocytes, FATP5, are protected from diet-induced NAFLD, indicating that excess fatty-acid transport into the hepatocyte is required for NAFLD and hepatic insulin resistance<sup>35</sup>.

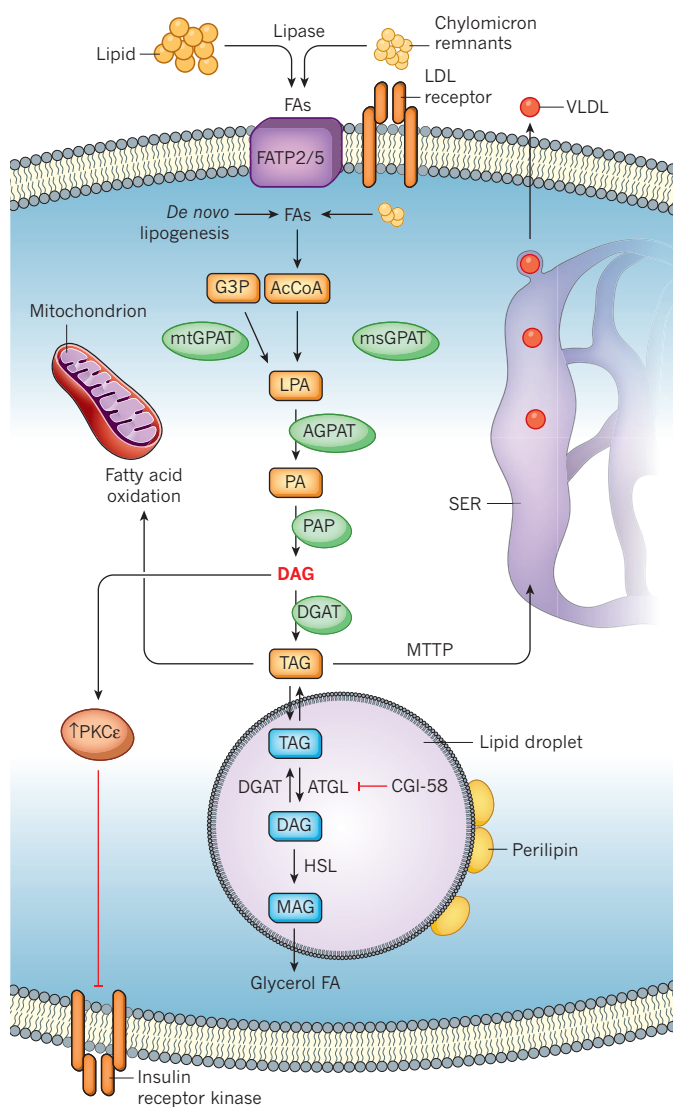
Studies in mice and humans have implicated adipose tissue lipolysis as an important source of fatty acids that promote NAFLD and hepatic insulin resistance. Whole-body lipolysis increases with total fat mass in humans<sup>36,37</sup>; however, the relationship between lipolysis and insulin sensitivity seems to be largely independent of body mass. For example, insulin-resistant obese adolescents have higher visceral fat content than their weight-matched, insulin-sensitive counterparts<sup>38</sup>. The relationship between adipose lipolysis and hepatic lipid content is exemplified by



**Figure 1 | Molecular mechanism by which excess diacylglycerol leads to hepatic insulin resistance and hyperglycaemia.** Increases in liver diacylglycerol (DAG) cause protein kinase C $\epsilon$  (PKC $\epsilon$ ) activation and translocation to the cell membrane, which results in inhibition of insulin signalling. Reduced phosphorylation of insulin receptor substrate-2 (IRS2) and PI(3)K impairs Akt2 activity by reductions in 3-phosphoinositide-dependent protein kinase 1 (PDK1) activity, suppressing glycogen synthase kinase-3 (GSK3) phosphorylation and reducing insulin-stimulated liver glycogen synthesis through reduced glycogen synthase (GS) activity. Impaired Akt2 activity also reduces insulin suppression of hepatic gluconeogenesis by promoting Forkhead box protein O1 (FOXO1) translocation to the nucleus due to reduced phosphorylation and increasing expression of the gluconeogenic proteins pyruvate carboxylase (PC), phosphoenolpyruvate carboxykinase (PEPCK), glucose-6-phosphatase (G6Pase). PIP3, phosphatidylinositol (3,4,5)-triphosphate.

manipulation of the genes that regulate adipose lipolysis. As in humans, the effect of lipolysis on insulin sensitivity in rodents seems to be independent of body weight. Obese mice lacking the fatty-acid-binding protein FABP in adipocytes are more insulin sensitive than their obese littermates with normal FABP<sup>39</sup>. Conversely, leptin-deficient obesity-prone mice with increased rates of adipocyte lipolysis due to knockout of the gene encoding adipocyte phospholipase A2 show increases in ectopic lipid storage and insulin resistance despite reduced body weight compared with littermates with normal lipolytic rates<sup>40</sup>. These data suggest a facilitative role for the increases in adipose tissue lipolysis in providing substrates for ectopic lipid deposition and insulin resistance.

These genetic rodent models inform our understanding of human disease. Patients with conditions resulting from mutations in LpL (for example, hyperlipoproteinaemia type 1) are prone to developing insulin resistance<sup>41</sup>. In humans, the single nucleotide polymorphism (SNP) rs56225452, putatively representing a gain-of-function mutation in the FATP5 promoter, was associated with insulin resistance and NAFLD<sup>42</sup>. In Asian Indian individuals, as well as those of other ethnic groups, variants (C-482T, T-455C or both) in apolipoprotein C3 (APOC3), which can inhibit LpL and hepatic lipase, are associated with hypertriglyceridaemia and NAFLD<sup>43</sup>. These polymorphisms led to around 30% higher plasma APOC3 concentrations and post-prandial hypertriglyceridaemia through the inhibitory effect of APOC3 on LpL activity. As a result, the livers of individuals with APOC3 variants take up a greater amount of lipid from chylomicrons, remnants of lipoprotein particles, predisposing these lean subjects to NAFLD and hepatic insulin resistance. These results



**Figure 2 | NAFLD develops due to an imbalance between lipid supply and demand.** Fatty acids (FAs) derived from lipolysis and from chylomicron remnants are taken up through fatty-acid transport proteins (FATPs), mainly FATP2 and FATP5 in the liver; chylomicron remnants are also taken up through the low-density lipoprotein (LDL) receptor. A small fraction of intracellular fatty acid supply in the liver also comes from *de novo* lipogenesis in the cytosol. Fatty acids can also be re-esterified to lysophosphatidic acid (LPA) by acyl-coenzyme A (AcCoA) and the conversion of glycerol 3-phosphate (G3P) by either mitochondrial glycerol-3-phosphate acyltransferase (mtGPAT) or microsomal GPAT (msGPAT). Fatty-acyl CoAs (shown here as phosphatidic acid, PA) formed by 1-acylglycerol-3-phosphate O-acyltransferase-2 (AGPAT2) are then added to the glycerol backbone by phosphatidic acid phosphatase (PAP) to generate diacylglycerol (DAG), and by diacylglycerol acyltransferases (DGAT) to generate triacylglycerol (TAG). Increased DAG causes protein kinase C $\epsilon$  (PKC $\epsilon$ ) translocation to the cell membrane, which inhibits insulin signalling. Lipids may also be sequestered in lipid droplets as monoacylglycerol (MAG), DAG and TAG, but these are not thought to be responsible for hepatic insulin resistance. By inhibition of adipose triglyceride lipase (ATGL), comparative gene identification-58 (CGI-58) bound to perilipin is mainly responsible for lipid sequestration in the droplet. By contrast, intracellular hepatic lipid content is reduced by two mechanisms: mitochondrial fatty acid oxidation and export from the smooth endoplasmic reticulum (SER) as very-low-density lipoprotein (VLDL). HSL, hormone-sensitive lipase; MTTP, microsomal triglyceride transfer protein.

were replicated in another cohort of lean males of European descent<sup>44</sup>. Of note, this *APOC3*-gene-environment interaction has only been observed in lean males, probably reflecting a protective effect of oestradiol on the

ability of *APOC3* to inhibit LpL and promote ectopic fat storage<sup>45</sup>, and the ability of obesity-associated NAFLD<sup>44</sup> to mask the relatively subtle affect of this gene-environment interaction on the development of hepatic insulin resistance<sup>46,47</sup>.

The effect of increased plasma Apoc3 concentrations on the development of NAFLD and hepatic insulin resistance has been genetically validated in transgenic mice that have increased hepatic overexpression of Apoc3. These mice are more prone to diet-induced NAFLD and diacylglycerol-PKC $\epsilon$ -induced hepatic insulin resistance than their wild-type littermates<sup>48</sup>. Interestingly, although hypertriglyceridaemia was present in transgenic mice fed both a normal and high-fat diet, severe hepatic steatosis and hepatic insulin resistance only developed in Apoc3 transgenic mice fed a high-fat diet, reflecting an important gene-environment interaction. Moreover, the phenotype was due to both inhibition of peripheral lipase activity and diminished hepatic triacylglycerol export as very-low-density lipoprotein (VLDL)<sup>48</sup>.

### Lessons learned from lipodystrophy

The importance of adipose tissue lipid storage is exemplified when adipose tissue is altogether absent. In 'fatless' mice expressing the dominant-negative protein A-ZIP/F-1 in adipocytes, the absence of visceral and peripheral fat leads to ectopic lipid accumulation and severe hepatic and muscle insulin resistance. Insulin resistance can be corrected by transplantation of white adipose tissue from normal mice, further illustrating the importance of adipose tissue as a 'safe' storage depot<sup>49</sup>. Lipotrophic mice with the gene encoding peroxisome proliferator-activated receptor- $\gamma$  (Ppar- $\gamma$ ) knocked out in white adipose tissue or with the gene encoding hepatic 1-acylglycerol-3-phosphate O-acyltransferase 2 (Agpat2) knocked out globally display a similar phenotype: the loss of visceral and subcutaneous fat is associated with hepatic steatosis and hepatic insulin resistance<sup>50,51</sup>.

Similar associations are evident in humans with lipodystrophies. Patients with these disorders represent a rare example of severe hepatic insulin associated with extreme hepatic steatosis in the absence of visceral or peripheral fat accumulation<sup>52-54</sup>. Leptin treatment can decrease calorie intake and effectively normalize hepatic lipid content and hepatic insulin action<sup>53,54</sup>. Similarly, patients with partial lipodystrophy owing to mutations in the scaffolding protein perilipin-1, which inhibits adipose triglyceride lipase, have reduced peripheral fat mass but develop NAFLD because of increased adipose tissue lipolysis resulting from inhibition of adipose tissue triglyceride lipase<sup>55</sup>. These patients can also develop profound insulin resistance. These data again point to ectopic lipid accumulation in the liver, which might occur as a result of the diversion of substrates from other storage depots, as the crucial mediator of hepatic insulin resistance.

### Regulation of hepatic lipid synthesis

Hepatic triacylglycerol synthesis is the sum of two main processes: the synthesis of fatty acids (*de novo* lipogenesis, DNL) and esterification of fatty acids into fatty-acid glyceride species (for example, mono-, di- and triacylglyceride).

#### Contributions of *de novo* lipogenesis to triacylglycerol synthesis

Although DNL is thought to make a relatively small contribution to hepatic triacylglycerol accumulation relative to esterification<sup>56,57</sup>, rates of postprandial DNL do increase significantly in both young and elderly patients with NAFLD<sup>58-60</sup>. Diet-induced NAFLD might stimulate a feed-forward loop exacerbating DNL and ectopic lipid deposition. For example, fructose-fed hamsters have hypertriglyceridaemia, NAFLD and insulin resistance associated with increased DNL<sup>61</sup>, and because fructose inhibits fatty-acid oxidation both directly and indirectly, excess fructose intake is likely to stimulate DNL and hepatic insulin resistance<sup>62,63</sup>. However, reducing lipogenic gene expression by knockdown of the upstream gene encoding peroxisome proliferator-activated receptor  $\gamma$  coactivator-1 $\beta$  (PGC-1 $\beta$ )<sup>64</sup> protects against fructose-induced hepatic insulin resistance<sup>65</sup>. Similarly knockdown of the genes encoding the acetyl

CoA carboxylases ACC1 and ACC2, which are crucial in the regulation of DNL and lipid oxidation, respectively, reduced liver triacylglycerol and diacylglycerol content, decreased PKC $\epsilon$  activation and protected mice from lipid-induced hepatic insulin resistance<sup>66</sup>.

### Skeletal muscle insulin resistance promotes hepatic lipogenesis

Skeletal muscle insulin resistance typically accompanies insulin resistance at other sites, possibly because of the diversion of substrates from insulin-resistant muscle to the liver (Fig. 3). The independent effect of muscle insulin resistance to exacerbate NAFLD has also been demonstrated in rodents. Mice that lack insulin-responsive glucose transporter 4 (Glut4) in muscle have NAFLD<sup>67</sup>. Similarly, mice lacking the Akt substrate As160 have decreased glucose uptake in adipose tissue and slow-twitch muscles, and develop insulin resistance in the liver as well as adipose tissue and skeletal muscle<sup>68</sup>. Muscle insulin resistance might also be independently associated with NAFLD. For example, the severity of NAFLD in mice fed a high-fat, high-cholesterol diet was demonstrated to correlate with peripheral insulin resistance, and not with hepatic insulin resistance<sup>69</sup>.

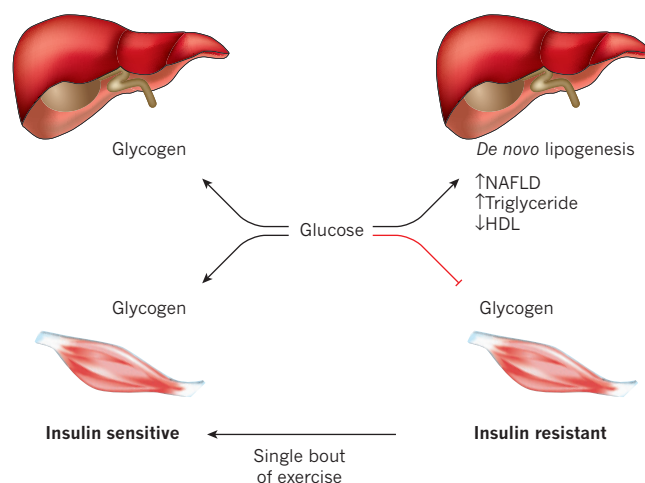
These results have been translated to humans, in which selective muscle insulin resistance in healthy young lean individuals has been shown to predispose them to increased hepatic DNL, hepatic triacylglycerol accumulation and atherogenic dyslipidaemia after eating high-carbohydrate meals. This is because ingested glucose is diverted away from muscle glycogen storage to the liver, in which it is converted to triacylglycerol driven by the compensatory hyperinsulinaemia that is secondary to muscle insulin resistance<sup>57</sup>. Further evidence to support this hypothesis stems from a study demonstrating that a single 45 minute bout of exercise on an elliptical trainer increased postprandial muscle glycogen synthesis following carbohydrate ingestion, resulting in a 40% reduction in hepatic DNL and a 30% reduction in hepatic triglyceride synthesis<sup>70</sup>.

### Fatty acid esterification contributes to triacylglycerol synthesis

Most liver triglyceride is formed through esterification of fatty acids<sup>56,57</sup>. Diacylglycerol is an intermediate in the esterification pathway and, thus, genetic models that manipulate hepatic lipid esterification can be used to further examine the diacylglycerol–novel-PKCs hypothesis of insulin resistance. Rats overexpressing the rate-controlling enzyme in triglyceride esterification, mitochondrial glycerol-3-phosphate acyltransferase 1 (GPAT1), have hepatic insulin resistance associated with increased PKC $\epsilon$  activity<sup>71</sup>. These data are in contrast to mice in which the gene encoding mitochondrial GPAT was knocked down. These animals exhibit suppression of PKC $\epsilon$  activity and improved hepatic insulin sensitivity<sup>72</sup>, offering further evidence in support of the diacylglycerol–PKC $\epsilon$  hypothesis of hepatic insulin resistance.

Diacylglycerol acyltransferase 2 (DGAT2) catalyses the final step in triglyceride synthesis from diacylglycerol. Although inhibition of DGAT2 may be expected to acutely increase cellular diacylglycerol content, chronic reduction in hepatic DGAT2 expression due to ASO treatment results in decreased hepatic diacylglycerol content due to downregulation of the lipogenic pathway. Consistent with the diacylglycerol–PKC $\epsilon$  hypothesis this reduction in hepatic diacylglycerol content was associated with reduced PKC $\epsilon$  activation and protection from lipid-mediated hepatic insulin resistance<sup>73,74</sup>.

The effects of hepatic overexpression of DGAT2 are less clear. Monetti *et al.* reported that although transgenic mice had increased hepatic diacylglycerol content, there was no impact on hepatic insulin sensitivity<sup>75</sup>. Jornayvaz *et al.* also demonstrated an increase in hepatic diacylglycerol content in the same mice, but reported increased PKC $\epsilon$  activation, decreased hepatic insulin signalling and hepatic insulin resistance<sup>76</sup>. Technical differences in study execution could explain this difference. In hyperinsulinaemic–euglycaemic clamp studies to test insulin resistance, both groups found that DGAT2 transgenic mice fed a normal diet failed to normally suppress hepatic glucose production. However, the teams reported differing results for wild-type mice fed a normal diet. Jornayvaz *et al.* showed that these control mice had normal insulin suppression of hepatic glucose production, whereas Monetti *et al.* reported that



**Figure 3 | Mechanism by which selective skeletal muscle insulin resistance contributes to hepatic insulin resistance.**

In insulin-sensitive subjects, insulin stimulates glycogen synthesis in both liver and muscle; however, in those with skeletal muscle insulin resistance, insulin fails to promote glycogen synthesis, diverting substrate to *de novo* lipogenesis. Increased lipid synthesis in patients with muscle insulin resistance thus produces non-alcoholic fatty liver disease (NAFLD), with increased triglyceride and reduced high-density lipoprotein (HDL) export from the liver. However, these defects in muscle insulin signalling can be reversed by a single 45 minute bout of exercise.

control mice did not suppress hepatic glucose production. Thus, the key to interpreting the phenotype attributed to increased hepatic DGAT2 expression and diacylglycerol accumulation in this model is whether or not the control animals fed a normal diet had complete suppression of hepatic glucose production in response to insulin. In a different study in humans, SNPs in DGAT2 predicted a smaller decrease in liver fat content after very modest (3 kg) weight loss compared with individuals without SNPs in DGAT2. However, insulin signalling was not measured in these studies, and the participants' very modest weight loss may prevent us from uncovering meaningful information about the role of DGAT2 (ref. 77). Further studies are needed to understand the role of DGAT2 in hepatic insulin sensitivity and its potential link to whole-body adiposity.

### Lipid export

Export of triglyceride as VLDL is the only means of reducing hepatic lipid content other than through fat oxidation. NAFLD-associated insulin resistance is seen in genetic models of impaired VLDL export<sup>78</sup>, whereas, in models of increased VLDL export, NAFLD is ameliorated, independent of body weight<sup>79,80</sup>. As previously discussed, impaired hepatic lipid export also contributes to the development of NAFLD and hepatic insulin resistance in mice overexpressing ApoC3 (ref. 48).

### Regulation of hepatic lipid oxidation

Rodent models with altered fat oxidation can be used to both test relationships between diacylglycerol, PKC $\epsilon$  activation and insulin resistance, and to validate potential therapeutic targets. Mice lacking Ppar- $\alpha$ , a key regulator of hepatic lipid oxidation, are prone to NAFLD and fail to benefit from the insulin sensitizing effects of omega-3 fatty acids<sup>81</sup>. Mice that are genetically deficient in the dehydrogenase LCAD have diminished mitochondrial fatty-acid oxidation and increased *de novo* diacylglycerol synthesis, increased liver PKC $\epsilon$  activity and hepatic insulin resistance<sup>82</sup>. Interestingly, humans and mice with loss-of-function mutations in LCAD exhibit fasting hypoglycaemia, which has been ascribed to defects in amino-acid metabolism and may not be related to alterations in hepatic insulin sensitivity<sup>83</sup>. Hypoglycaemia has also been dissociated from insulin resistance in mice fed a ketogenic diet. These mice develop profound hepatic insulin resistance associated with increased diacylglycerol concentrations and PKC $\epsilon$  activation, despite hypoglycaemia stemming from reductions in fasting hepatic gluconeogenesis<sup>84</sup>.



Changes in whole-body energy metabolism also affect hepatic lipid balance. Thyroid hormone receptor- $\alpha$  knockout mice have increased energy expenditure (as well as decreased expression of lipogenic enzymes) and are protected from diet-induced NAFLD, PKC $\epsilon$  activation and hepatic insulin resistance<sup>85</sup>. Similarly, infusion of fibroblast growth factor 21 (Fgf21) in mice resulted in increased energy expenditure in the liver and white adipose tissue, lowered liver diacylglycerol concentrations, decreased hepatic PKC $\epsilon$  translocation and protection from lipid-induced hepatic insulin resistance<sup>86</sup>. By contrast, mice deficient for fatty acid amide hydrolase (FAAH) have reduced whole-body energy expenditure due to diminished hypothalamic–pituitary axis activity and impaired thyroid function. The absence of FAAH resulted in increased liver diacylglycerol, increased PKC $\epsilon$  translocation and hepatic insulin resistance<sup>87</sup>. These data again indicate that the link between NAFLD and insulin resistance is an increase in diacylglycerol concentration leading to activation of PKC $\epsilon$  activity.

As already mentioned, inhibition of ACC1 and ACC2 leads to both a decrease in lipid synthesis and an increase in lipid oxidation. The latter effect is due to the disinhibition of carnitine palmitoyltransferase-1 (CPT1) and more long-chain fatty acyl CoAs entering the mitochondria. Mutation of key serine residues in Acc1 and Acc2 in mice prevents inactivation of these enzymes by AMP-activated protein kinase (AMPK), leading to increased lipid synthesis and decreased lipid oxidation<sup>88</sup>. These mice have increased hepatic diacylglycerol content, activation of PKC $\epsilon$  and develop hepatic insulin resistance. By contrast, AMPK is activated in mice with a deletion of the mitochondrial gene encoding the sodium-dicarboxylate cotransporter Slc13a5, the mammalian homologue of the INDY protein in *Drosophila*. The increased AMPK activity leads to reductions in Acc1 and Acc2 activity, resulting in decreased liver diacylglycerol content and reduced PKC $\epsilon$  translocation. Together, these changes protected *Slc13a5* knockout mice from both diet- and age-associated hepatic insulin resistance<sup>89</sup>.

Many groups have studied whether people with NAFLD have alterations in hepatic oxidative flux either as a cause or consequence of ectopic lipid accumulation. Results in humans have been mixed. When an indirect tracer method was used to assess hepatic tricarboxylic acid (TCA) cycle flux and anaplerosis, both were found to be markedly increased in individuals with NAFLD<sup>90,91</sup>. But when a <sup>31</sup>P magnetic resonance spectroscopy technique was used to assess hepatic ATP production, a reduction in hepatic energy metabolism was found in individuals with type 2 diabetes<sup>92,93</sup>. Because the role of hepatic oxidative flux is key to both the understanding of pathogenesis and potential treatment of NAFLD and type 2 diabetes, the development of methods to directly measure rates hepatic oxidative metabolism is of crucial importance<sup>94</sup>.

### Dissociation of NAFLD and hepatic insulin resistance

The requirement of NAFLD for hepatic insulin resistance has been questioned by several groups and has been recently reviewed<sup>95–97</sup>. It is well established that it is possible to experimentally induce insulin resistance without NAFLD, or induce NAFLD without insulin resistance under certain conditions. For instance, blocking hepatic VLDL secretion with a choline-deficient diet or by genetic modification of the export machinery increases hepatic triglyceride concentrations but does not cause insulin resistance<sup>98,99</sup>. Similarly, mice with liver-specific knockout of the gene encoding the protein phosphatase Shp1 develop NAFLD, but are protected from insulin resistance<sup>100</sup>. Hepatic carbohydrate responsive element-binding protein and sterol regulatory element-binding protein-1 have recently been independently implicated in dissociating NAFLD and insulin resistance in mice and humans<sup>101,102</sup>. In addition, mice with the gene encoding the lipase activator CGI-58 knocked down have profound hepatic steatosis due to suppression of hepatic triglyceride lipolysis. However, these mice are protected against high-fat-diet-induced hepatic insulin resistance<sup>103</sup>. To explain this paradox, the authors of one study examined the subcellular localization of diacylglycerols in liver cells and found that knockdown of the gene encoding CGI-58 promoted diacylglycerol accumulation in lipid droplets, while protecting

against diacylglycerol accumulation in the cell membrane and preventing PKC $\epsilon$  translocation to the cell membrane<sup>104</sup>. These results are consistent with earlier studies in humans, also showing that compartmentation of diacylglycerols in the cytosolic<sup>29</sup> and membrane<sup>105</sup> compartments could be an important factor in the pathogenesis of liver and muscle insulin resistance. The different results between diacylglycerol accumulation in the membrane and the cytosolic compartments might reflect differences in the measurement and fractionating or lipid-extraction techniques used in the different studies and/or in the length of fasting. Nevertheless, the CGI-58 ASO data clearly indicate that lipids sequestered in lipid droplets do not promote PKC $\epsilon$  activation and hepatic insulin resistance, and that this compartmentation of diacylglycerols and triacylglycerols in lipid droplets may explain other models of NAFLD that are not associated with hepatic insulin resistance, as discussed later. Future work will better discern the importance of specific lipid compartments in the pathogenesis of insulin resistance.

Similarly, liver-specific *Hdac3* knockout mice are prone to developing hepatic steatosis when fed both normal and high-fat diets and have been shown to be protected from lipid-induced hepatic resistance<sup>106</sup>. This apparent disconnect is probably also due to altered partitioning of diacylglycerols to the lipid droplet. Loss of *Hdac3* in mice was associated with an increase in the lipid droplet protein Plin2 and decreased activation of PKC $\epsilon$ . When Plin2 expression was normalized with a specific ASO, the improvements in glucose tolerance were no longer evident. Taken together these data strongly suggest that diacylglycerols need access to a particular subcellular compartment to inhibit IRTK activity, whereas diacylglycerol accumulation in the lipid droplet is protective and does not lead to inhibition of IRTK activity and insulin resistance.

Genetic variants in the triglyceride hydrolase PNPLA3 (also known as adiponutrin) have been proposed to disassociate NAFLD from insulin resistance in normal weight and obese humans<sup>46</sup>. However, this conclusion was based on fasting plasma glucose and insulin measurements, and hepatic insulin responsiveness was not directly assessed in this study. Furthermore, because many of the subjects in this study were obese and relatively insulin resistant it is difficult to determine if PNPLA3 could exacerbate hepatic insulin resistance in these individuals. Additional insights into the role of PNPLA3 have been gained using an ASO to decrease expression of PNPLA3 in rats<sup>107</sup>. Reduced expression of PNPLA3 decreased hepatic lipid content owing to decreased lipid esterification, pointing to the potential duality of PNPLA3's functions in both triglyceride synthesis and hydrolysis<sup>108</sup>. Similarly, in obese Taiwanese children, variants (which had previously been identified<sup>109</sup> to strongly correlate with NAFLD in adults of European ancestry who were not all obese) did not correlate with NAFLD, but again this may be due to the obfuscating effects of obesity in this cohort; however, there was a significant association between glucokinase regulatory protein (GCKR) and NAFLD<sup>110</sup>.

### Reversal of NAFLD ameliorates hepatic insulin resistance

The most effective intervention to reverse NAFLD and hepatic insulin resistance in humans is weight loss. Hepatic steatosis quickly resolves in both obese patients with type 2 diabetes and non-diabetic normal weight individuals with NAFLD after a hypocaloric diet and a modest weight loss of less than 10% of total body weight<sup>111</sup>. This is accompanied by resolution of hepatic steatosis and normalization of fasting plasma glucose concentrations, hepatic glucose production and hepatic insulin sensitivity<sup>111,112</sup>. However, recidivism following weight loss is extremely common: less than 50% of those who have lost more than 10% of their body weight are able to maintain the weight loss after 1 year<sup>113</sup>, and after 5 years, less than 25% have maintained their weight<sup>114</sup>. Nevertheless, improvements in insulin sensitivity and NAFLD can persist for 2 years after weight loss in overweight and obese individuals even after the weight is regained, arguing that there is a dissociation between obesity and insulin sensitivity or NAFLD<sup>115</sup>.

In regards to potential medical therapies to treat NAFLD,

thiazolidinediones, which are potent PPAR- $\gamma$  activators, have been shown to lead to significant reductions in liver-fat content and improvements in hepatic insulin sensitivity<sup>116–120</sup>. This effect probably occurs by thiazolidinedione activation of PPAR- $\gamma$  in subcutaneous fat tissue leading to increased insulin sensitivity and suppression of lipolysis, resulting in a redistribution of liver fat to the subcutaneous fat cells<sup>116</sup>.

Increasing hepatic mitochondrial fat oxidation by promoting subtle increases in mitochondrial uncoupling could be another therapeutic target for NAFLD, hepatic insulin resistance and type 2 diabetes. Promoting mitochondrial uncoupling by treating rats fed a high-fat diet with low doses of the mitochondrial protonophore 2,4-dinitrophenol (DNP) was shown to reduce hepatic triglyceride content, increase insulin-stimulated IRS2 tyrosine phosphorylation and PI(3)K activity, and improve insulin-mediated suppression of hepatic glucose production<sup>25</sup>. Consistent with these findings, simply reversing hepatic steatosis using a liver-targeted DNP to increase hepatic mitochondrial fat oxidation by about 60% was found to reverse hypertriglyceridaemia, hepatic and peripheral insulin resistance, and hyperglycaemia in rat models of NAFLD and type 2 diabetes. Furthermore, correction of liver and muscle insulin resistance was associated with marked reductions in diacylglycerol-PKC $\epsilon$  and diacylglycerol-PKC $\theta$  activity in liver and muscle, respectively; and occurred independently of any changes in body weight, inflammatory mediators, FGF21, adiponectin or liver ceramide content<sup>121</sup>. Moreover, liver-targeting DNP resulted in a 50-fold increase in the ratio of toxic to effective doses compared with DNP. Taken together these data support the key role of diacylglycerol activation of novel PKC in mediating liver insulin resistance and demonstrate the potential feasibility of dissociating the toxic effects of DNP from its beneficial effects to promote subtle increases in hepatic mitochondrial uncoupling and hepatic fat oxidation to treat the related epidemics of NAFLD and type 2 diabetes.

## Outlook

Hepatic insulin resistance is a complex phenomenon. Although many questions regarding the nature of the insulin-signalling defect remain unanswered, hepatic insulin resistance is almost universally associated with increases in hepatic triacylglycerol and diacylglycerol concentrations, with the latter leading to PKC $\epsilon$  activation and subsequent inhibition of IRTK activity. The diacylglycerol-PKC $\epsilon$  hypothesis of hepatic insulin resistance has recently been validated in humans with NAFLD. Exceptions to this rule can be attributed to compartmentation of diacylglycerols to the lipid droplet, in which it does not lead to inhibition of IRTK activity. NAFLD occurs when lipid supply to the liver exceeds rates of lipid oxidation and lipid export. Thus therapies targeted to reduce fatty-acid delivery to the liver, suppress diacylglycerol production, or raise mitochondrial fat oxidation by promoting subtle increases in hepatic mitochondrial uncoupling are of great interest to ameliorate NAFLD and type 2 diabetes. ■

Received 26 February; accepted 14 March 2014.

- Browning, J. D. *et al.* Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology* **40**, 1387–1395 (2004).
- Smits, M. M., Ioannou, G. N., Boyko, E. J. & Utzschneider, K. M. Non-alcoholic fatty liver disease as an independent manifestation of the metabolic syndrome: results of a US national survey in three ethnic groups. *J. Gastroenterol. Hepatol.* **28**, 664–670 (2013).
- Fan, J.-G. *et al.* Prevalence of and risk factors for fatty liver in a general population of Shanghai, China. *J. Hepatol.* **43**, 508–514 (2005).
- Amarapurkar, D. N. *et al.* How common is non-alcoholic fatty liver disease in the Asia-Pacific region and are there local differences? *J. Gastroenterol. Hepatol.* **22**, 788–793 (2007).
- Petersen, K. F. *et al.* Increased prevalence of insulin resistance and nonalcoholic fatty liver disease in Asian-Indian men. *Proc. Natl Acad. Sci. USA* **103**, 18273–18277 (2006).  
**This study reported ethnic differences in the prevalence of NAFLD and insulin resistance.**
- Das, K. *et al.* Nonobese population in a developing country has a high prevalence of nonalcoholic fatty liver and significant liver disease. *Hepatology* **51**, 1593–1602 (2010).
- Tolman, K. G., Fonseca, V., Dalpiaz, A. & Tan, M. H. Spectrum of liver disease in type 2 diabetes and management of patients with diabetes and liver disease. *Diabetes Care* **30**, 734–743 (2007).

- Wanless, I. R. & Lentz, J. S. Fatty liver hepatitis (steatohepatitis) and obesity: an autopsy study with analysis of risk factors. *Hepatology* **12**, 1106–1110 (1990).  
**This study reported that steatohepatitis was sevenfold more common in severely obese compared with patients of normal weight contributing to type 2 diabetes risk.**
- Silverman, J. F. *et al.* Liver pathology in morbidly obese patients with and without diabetes. *Am. J. Gastroenterol.* **85**, 1349–1355 (1990).
- Fabbri, E., Sullivan, S. & Klein, S. Obesity and nonalcoholic fatty liver disease: biochemical, metabolic, and clinical implications. *Hepatology* **51**, 679–689 (2010).  
**This review summarizes the link between obesity, NAFLD and insulin resistance and the possible role of inflammation in these processes.**
- Shulman, G. I. Cellular mechanisms of insulin resistance. *J. Clin. Invest.* **106**, 171–176 (2000).  
**This review describes the cellular and molecular mechanisms of liver and muscle insulin resistance and proposes the diacylglycerol and novel PKC hypothesis of lipid-induced insulin resistance.**
- Kleiner, D. E. *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**, 1313–1321 (2005).
- Hui, J. M. *et al.* Long-term outcomes of cirrhosis in nonalcoholic steatohepatitis compared with hepatitis C. *Hepatology* **38**, 420–427 (2003).
- Ratziu, V., Bellentani, S., Cortez-Pinto, H., Day, C. & Marchesini, G. A position statement on NAFLD/NASH based on the EASL 2009 special conference. *J. Hepatol.* **53**, 372–384 (2010).
- Charlton, M. R. *et al.* Frequency and outcomes of liver transplantation for nonalcoholic steatohepatitis in the United States. *Gastroenterology* **141**, 1249–1253 (2011).
- Samuel, V. T. & Shulman, G. I. Mechanisms for insulin resistance: common threads and missing links. *Cell* **148**, 852–871 (2012).  
**This review provides a balanced and detailed discussion of the potential roles of inflammation, ER stress, ceramides and other factors in the pathogenesis of liver and muscle insulin resistance.**
- Cheng, Z., Tseng, Y. & White, M. F. Insulin signaling meets mitochondria in metabolism. *Trends Endocrinol. Metab.* **21**, 589–598 (2010).
- Hanke, S. & Mann, M. The phosphotyrosine interactome of the insulin receptor family and its substrates IRS-1 and IRS-2. *Mol. Cell. Proteomics* **8**, 519–534 (2009).
- Franke, T. F., Kaplan, D. R., Cantley, L. C. & Toker, A. Direct regulation of the Akt proto-oncogene product by phosphatidylinositol-3,4-bisphosphate. *Science* **275**, 665–668 (1997).
- Adina-Zada, A. *et al.* Allosteric regulation of the biotin-dependent enzyme pyruvate carboxylase by acetyl-CoA. *Biochem. Soc. Trans.* **40**, 567–572 (2012).
- Adina-Zada, A., Zeczycki, T. N. & Attwood, P. V. Regulation of the structure and activity of pyruvate carboxylase by acetyl CoA. *Arch. Biochem. Biophys.* **519**, 118–130 (2012).
- Pilkis, S. J., el-Maghrabi, M. R. & Claus, T. H. Fructose-2,6-bisphosphate in control of hepatic gluconeogenesis. From metabolites to molecular genetics. *Diabetes Care* **13**, 582–599 (1990).
- Petersen, K. F., Laurent, D., Rothman, D. L., Cline, G. W. & Shulman, G. I. Mechanism by which glucose and insulin inhibit net hepatic glycogenolysis in humans. *J. Clin. Invest.* **101**, 1203–1209 (1998).
- Roden, M. *et al.* Mechanism of free fatty acid-induced insulin resistance in humans. *J. Clin. Invest.* **97**, 2859–2865 (1996).
- Samuel, V. T. *et al.* Mechanism of hepatic insulin resistance in non-alcoholic fatty liver disease. *J. Biol. Chem.* **279**, 32345–32353 (2004).  
**The authors of this paper established a model of selective hepatic insulin resistance and demonstrated that this resistance was associated with increased hepatic diacylglycerol content and increased PKC $\epsilon$  activation for the first time.**
- Samuel, V. T. *et al.* Inhibition of protein kinase C $\epsilon$  prevents hepatic insulin resistance in nonalcoholic fatty liver disease. *J. Clin. Invest.* **117**, 739–745 (2007).  
**This paper conclusively demonstrated the key role of PKC $\epsilon$  activation in mediating lipid-induced hepatic insulin resistance.**
- Dries, D. R., Gallegos, L. L. & Newton, A. C. A single residue in the C1 domain sensitizes novel protein kinase C isoforms to cellular diacylglycerol production. *J. Biol. Chem.* **282**, 826–830 (2007).
- Raddatz, K. *et al.* Time-dependent effects of Prkce deletion on glucose homeostasis and hepatic lipid metabolism on dietary lipid oversupply in mice. *Diabetologia* **54**, 1447–1456 (2011).
- Kumashiro, N. *et al.* Cellular mechanism of insulin resistance in nonalcoholic fatty liver disease. *Proc. Natl Acad. Sci. USA* **108**, 16381–16385 (2011).  
**This paper reports that intracellular diacylglycerol, associated with PKC $\epsilon$  activation, is the strongest predictor of insulin resistance in obese patients.**
- Magkos, F. *et al.* Intrahepatic diacylglycerol content is associated with hepatic insulin resistance in obese subjects. *Gastroenterology* **142**, 1444–1446.e2 (2012).
- Funke, A. *et al.* Cholesterol-induced hepatic inflammation does not contribute to the development of insulin resistance in male LDL receptor knockout mice. *Atherosclerosis* **232**, 390–396 (2014).
- Brown, M. S. & Goldstein, J. L. Selective versus total insulin resistance: a pathogenic paradox. *Cell Metab.* **7**, 95–96 (2008).
- Chavez, J. A. & Summers, S. A. Lipid oversupply, selective insulin resistance, and lipotoxicity: molecular mechanisms. *Biochim. Biophys. Acta* **1801**, 252–265 (2010).



34. Kim, J. K. *et al.* Tissue-specific overexpression of lipoprotein lipase causes tissue-specific insulin resistance. *Proc. Natl Acad. Sci. USA* **98**, 7522–7527 (2001).  
**This paper reports that overexpression of LpL in liver resulted in liver-specific triglyceride accumulation and liver-specific insulin resistance, whereas muscle-specific overexpression of LpL resulted in muscle-specific triglyceride accumulation and muscle-specific insulin resistance.**
35. Doerge, H. *et al.* Silencing of hepatic fatty acid transporter protein 5 *in vivo* reverses diet-induced non-alcoholic fatty liver disease and improves hyperglycemia. *J. Biol. Chem.* **283**, 22186–22192 (2008).
36. Mittendorfer, B., Magkos, F., Fabbri, E., Mohammed, B. S. & Klein, S. Relationship between body fat mass and free fatty acid kinetics in men and women. *Obesity (Silver Spring)* **17**, 1872–1877 (2009).
37. Pardini, E. *et al.* Increased expression and activity of hepatic lipase in the liver of morbidly obese adult patients in relation to lipid content. *Obes. Surg.* **19**, 894–904 (2009).
38. Weiss, R. *et al.* The 'obese insulin-sensitive' adolescent: importance of adiponectin and lipid partitioning. *J. Clin. Endocrinol. Metab.* **90**, 3731–3737 (2005).
39. Cao, H. *et al.* Regulation of metabolic responses by adipocyte/macrophage Fatty Acid-binding proteins in leptin-deficient mice. *Diabetes* **55**, 1915–1922 (2006).
40. Jaworski, K. *et al.* AdPLA ablation increases lipolysis and prevents obesity induced by high-fat feeding or leptin deficiency. *Nature Med.* **15**, 159–168 (2009).
41. Mingrone, G. *et al.* Triglyceride-induced diabetes associated with familial lipoprotein lipase deficiency. *Diabetes* **48**, 1258–1263 (1999).
42. Auinger, A. *et al.* A promoter polymorphism in the liver-specific fatty acid transport protein 5 is associated with features of the metabolic syndrome and steatosis. *Horm. Metab. Res.* **42**, 854–859 (2010).
43. Petersen, K. F. *et al.* Apolipoprotein C3 gene variants in nonalcoholic fatty liver disease. *N. Engl. J. Med.* **362**, 1082–1089 (2010).  
**The authors of this paper show that variants in APOC3 were associated with a high prevalence of NAFLD and insulin resistance in lean Asian-Indian men.**
44. Peter, A., Kantartzis, K., Machicao, F. & Machann, J. Visceral obesity modulates the impact of apolipoprotein C3 gene variants on liver fat content. *J. Obes.* **36**, 774–782 (2012).  
**This paper, which follows up on the findings of ref. 43 shows that the association between APOC3 is only observable in lean, not obese individuals, demonstrating that obesity may mask the predisposing effects of APOC3 genetic variants on NAFLD and insulin resistance.**
45. Camporez, J. P. G. *et al.* Cellular mechanism by which estradiol protects female ovariectomized mice from high-fat diet-induced hepatic and muscle insulin resistance. *Endocrinology* **154**, 1021–1028 (2013).
46. Verrijken, A., Beckers, S., Franque, S. & Hilden, H. A gene variant of PNPLA3, but not of APOC3, is associated with histological parameters of NAFLD in an obese population. *Obesity (Silver Spring)* **21**, 2138–2145 (2013).
47. Kozlitina, J., Boerwinkle, E., Cohen, J. C. & Hobbs, H. H. Dissociation between APOC3 variants, hepatic triglyceride content and insulin resistance. *Hepatology* **53**, 467–474 (2011).
48. Lee, H.-Y. *et al.* Apolipoprotein CIII overexpressing mice are predisposed to diet-induced hepatic steatosis and hepatic insulin resistance. *Hepatology* **54**, 1650–1660 (2011).  
**This study demonstrates that transgenic mice with hepatic overexpression of human APOC3 predisposes them to severe hepatic steatosis and hepatic insulin resistance when fed a high-fat diet, whereas there is no metabolic phenotype when they are fed a normal diet.**
49. Kim, J. K., Gavrilova, O., Chen, Y., Reitman, M. L. & Shulman, G. I. Mechanism of insulin resistance in A-ZIP/F-1 fatless mice. *J. Biol. Chem.* **275**, 8456–8460 (2000).  
**This study clearly illustrated the mechanism by which lipodystrophy syndromes lead to insulin resistance.**
50. Wang, F., Mullican, S. E., DiSpirito, J. R., Peed, L. C. & Lazar, M. A. Lipodystrophy and severe metabolic disturbance in mice with fat-specific deletion of PPAR $\gamma$ . *Proc. Natl Acad. Sci. USA* **110**, 18656–18661 (2013).
51. Cortés, V. A. *et al.* Leptin ameliorates insulin resistance and hepatic steatosis in Agpat2<sup>-/-</sup> lipodystrophic mice independent of hepatocyte leptin receptors. *J. Lipid Res.* **55**, 276–288 (2014).
52. Savage, D. B., Murgatroyd, P. R., Chatterjee, V. K. & O'Rahilly, S. Energy expenditure and adaptive responses to an acute hypercaloric fat load in humans with lipodystrophy. *J. Clin. Endocrinol. Metab.* **90**, 1446–1452 (2005).
53. Petersen, K. F. *et al.* Leptin reverses insulin resistance and hepatic steatosis in patients with severe lipodystrophy. *J. Clin. Invest.* **109**, 1345–1350 (2002).  
**This study established the mechanism by which leptin replacement therapy reverses liver and muscle insulin resistance in patients with lipodystrophy.**
54. Simha, V., Szczepaniak, L. S., Wagner, A. J., DePaoli, A. M. & Garg, A. Effect of leptin replacement on intrahepatic and intramyocellular lipid content in patients with generalized lipodystrophy. *Diabetes Care* **26**, 30–35 (2003).
55. Gandotra, S. *et al.* Perilipin deficiency and autosomal dominant partial lipodystrophy. *N. Engl. J. Med.* **364**, 740–748 (2011).
56. Diraison, F., Moulin, P. & Beylot, M. Contribution of hepatic *de novo* lipogenesis and reesterification of plasma non esterified fatty acids to plasma triglyceride synthesis during non-alcoholic fatty liver disease. *Diabetes Metab.* **29**, 478–485 (2003).
57. Petersen, K. F. *et al.* The role of skeletal muscle insulin resistance in the pathogenesis of the metabolic syndrome. *Proc. Natl Acad. Sci. USA* **104**, 12587–12594 (2007).
- In this paper the authors demonstrate that selective insulin resistance in skeletal muscle promotes the development of atherogenic dyslipidaemia and NAFLD.**
58. Donnelly, K. L. *et al.* Sources of fatty acids stored in liver and secreted via lipoproteins in patients with nonalcoholic fatty liver disease. *J. Clin. Invest.* **115**, 1343–1351 (2005).
59. Petersen, K. F. *et al.* Reversal of muscle insulin resistance by weight reduction in young, lean, insulin-resistant offspring of parents with type 2 diabetes. *Proc. Natl Acad. Sci. USA* **109**, 8236–8240 (2012).  
**This study provides evidence that skeletal muscle insulin resistance in young lean insulin-resistant offspring of parents with type 2 diabetes can be attributed to increased intramyocellular lipid content.**
60. Flannery, C., Dufour, S., Rabøl, R., Shulman, G. I. & Petersen, K. F. Skeletal muscle insulin resistance promotes increased hepatic *de novo* lipogenesis, hyperlipidemia, and hepatic steatosis in the elderly. *Diabetes* **61**, 2711–2717 (2012).
61. Bilz, S. *et al.* Activation of the farnesoid X receptor improves lipid metabolism in combined hyperlipidemic hamsters. *Am. J. Physiol. Endocrinol. Metab.* **290**, E716–E722 (2006).
62. Delarue, J., Normand, S., Couet, C. & Pachiaudi, C. Effects of free fatty acids on the metabolic response to oral fructose in lean healthy humans. *Int. J. Obes. Relat. Metab. Disord.* **20**, 130–136 (1996).
63. Zhang, C. *et al.* Endoplasmic reticulum stress is involved in hepatic SREBP-1c activation and lipid accumulation in fructose-fed mice. *Toxicol. Lett.* **212**, 229–240 (2012).
64. Lin, J. *et al.* Hyperlipidemic effects of dietary saturated fats mediated through PGC-1 $\beta$  coactivation of SREBP. *Cell* **120**, 261–273 (2005).
65. Nagai, Y. *et al.* The role of peroxisome proliferator-activated receptor  $\gamma$  coactivator-1  $\beta$  in the pathogenesis of fructose-induced insulin resistance. *Cell Metab.* **9**, 252–264 (2009).
66. Savage, D. B. *et al.* Reversal of diet-induced hepatic steatosis and hepatic insulin resistance by antisense oligonucleotide inhibitors of acetyl-CoA carboxylases 1 and 2. *J. Clin. Invest.* **116**, 817–824 (2006).  
**The authors of this study reported that decreasing hepatic expression of hepatic ACC1 and ACC2 in rats by ASO decreased hepatic lipogenesis and increased liver fat oxidation, resulting in protection from lipid-induced hepatic steatosis and hepatic insulin resistance.**
67. Kim, J. K. *et al.* Glucose toxicity and the development of diabetes in mice with muscle-specific inactivation of GLUT4. *J. Clin. Invest.* **108**, 153–160 (2001).
68. Wang, H. Y. *et al.* AS160 deficiency causes whole-body insulin resistance via composite effects in multiple tissues. *Biochem. J.* **449**, 479–489 (2013).
69. Asai, A. *et al.* Dissociation of hepatic insulin resistance from susceptibility of non-alcoholic fatty liver disease induced by a high fat and high carbohydrate diet in mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* <http://dx.doi.org/10.1152/ajpgi.00291.2013> (2014).
70. Rabøl, R., Petersen, K. F., Dufour, S., Flannery, C. & Shulman, G. I. Reversal of muscle insulin resistance with exercise reduces postprandial hepatic *de novo* lipogenesis in insulin resistant individuals. *Proc. Natl Acad. Sci. USA* **108**, 13705–13709 (2011).  
**The authors of this paper demonstrate that reversal of muscle insulin resistance in healthy young lean insulin-resistant individuals with a single 45 minute bout of elliptical exercise reversed the abnormal pattern of energy distribution of energy storage following carbohydrate ingestion, thus offering strong evidence in support of a key role for selective muscle insulin resistance in promoting NAFLD and atherogenic dyslipidaemia as proposed in ref. 57.**
71. Nagle, C. A. *et al.* Hepatic overexpression of glycerol-sn-3-phosphate acyltransferase 1 in rats causes insulin resistance. *J. Biol. Chem.* **282**, 14807–14815 (2007).
72. Neschen, S., Morino, K., Hammond, L. E., Zhang, D. & Liu, X. X. Prevention of hepatic steatosis and hepatic insulin resistance in mitochondrial acyl-CoA: glycerol-sn-3-phosphate acyltransferase 1 knockout mice. *Cell Metab.* **2**, 55–65 (2005).
73. Yu, X. X. *et al.* Antisense oligonucleotide reduction of DGAT2 expression improves hepatic steatosis and hyperlipidemia in obese mice. *Hepatology* **42**, 362–371 (2005).
74. Choi, C. S. *et al.* Suppression of diacylglycerol acyltransferase-2 (DGAT2), but not DGAT1, with antisense oligonucleotides reverses diet-induced hepatic steatosis and insulin resistance. *J. Biol. Chem.* **282**, 22678–22688 (2007).
75. Monetti, M. *et al.* Dissociation of hepatic steatosis and insulin resistance in mice overexpressing DGAT in the liver. *Cell Metab.* **6**, 69–78 (2007).
76. Jornayvaz, F. R. *et al.* Hepatic insulin resistance in mice with hepatic overexpression of diacylglycerol acyltransferase 2. *Proc. Natl Acad. Sci. USA* **108**, 5748–5752 (2011).
77. Kantartzis, K. *et al.* The DGAT2 gene is a candidate for the dissociation between fatty liver and insulin resistance in humans. *Clin. Sci.* **116**, 531–537 (2009).
78. Shindo, N. *et al.* Involvement of microsomal triglyceride transfer protein in nonalcoholic steatohepatitis in novel spontaneous mouse model. *J. Hepatol.* **52**, 903–912 (2010).
79. Morán-Ramos, S. *et al.* *Opuntia ficus indica* (nopal) attenuates hepatic steatosis and oxidative stress in obese Zucker (fa/fa) rats. *J. Nutr.* **142**, 1956–1963 (2012).
80. Singhal, N. S., Patel, R. T., Qi, Y., Lee, Y. S. & Ahima, R. S. Loss of resistin ameliorates hyperlipidemia and hepatic steatosis in leptin-deficient mice. *Am. J. Physiol. Endocrinol. Metab.* **295**, E331–E338 (2008).
81. Neschen, S. *et al.* n-3 Fatty acids preserve insulin sensitivity *in vivo* in a peroxisome proliferator-activated receptor- $\alpha$ -dependent manner. *Diabetes* **56**, 1034–1041 (2007).



82. Zhang, D. *et al.* Mitochondrial dysfunction due to long-chain Acyl-CoA dehydrogenase deficiency causes hepatic steatosis and hepatic insulin resistance. *Proc. Natl Acad. Sci. USA* **104**, 17075–17080 (2007).
83. Houten, S. M. *et al.* Impaired amino acid metabolism contributes to fasting-induced hypoglycemia in fatty acid oxidation defects. *Hum. Mol. Genet.* **22**, 5249–5261 (2013).
84. Jornayvaz, F. R. *et al.* A high fat, ketogenic diet causes hepatic insulin resistance in mice, despite increasing energy expenditure and preventing weight gain. *Am. J. Physiol. Endocrinol. Metab.* **299**, E808–E815 (2010).
85. Jornayvaz, F. R. *et al.* Thyroid hormone receptor- $\alpha$  gene knockout mice are protected from diet-induced hepatic insulin resistance. *Endocrinology* **153**, 583–591 (2012).
86. Camporez, J. P. G. *et al.* Cellular mechanisms by which FGF21 improves insulin sensitivity in male mice. *Endocrinology* **154**, 3099–3109 (2013).
87. Brown, W. H. *et al.* Fatty acid amide hydrolase ablation promotes ectopic lipid storage and insulin resistance due to centrally mediated hypothyroidism. *Proc. Natl Acad. Sci. USA* **109**, 14966–14971 (2012).
88. Fullerton, M. D. *et al.* Single phosphorylation sites in Acc1 and Acc2 regulate lipid homeostasis and the insulin-sensitizing effects of metformin. *Nature Med.* **19**, 1649–1654 (2013).
89. Birkenfeld, A. L. *et al.* Deletion of the mammalian INDY homolog mimics aspects of dietary restriction and protects against adiposity and insulin resistance in mice. *Cell Metab.* **14**, 184–195 (2011).
90. Sunny, N. E., Parks, E. J., Browning, J. D. & Burgess, S. C. Excessive hepatic mitochondrial TCA cycle and gluconeogenesis in humans with nonalcoholic fatty liver disease. *Cell Metab.* **14**, 804–810 (2011).
91. Satapati, S. *et al.* Elevated TCA cycle function in the pathology of diet-induced hepatic insulin resistance and fatty liver. *J. Lipid Res.* **53**, 1080–1092 (2012).
92. Szendroedi, J. *et al.* Abnormal hepatic energy homeostasis in type 2 diabetes. *Hepatology* **50**, 1079–1086 (2009).
93. Schmid, A. I. *et al.* Liver ATP synthesis is lower and relates to insulin sensitivity in patients with type 2 diabetes. *Diabetes Care* **34**, 448–453 (2011).
94. Befroy, D. E. *et al.* Direct assessment of hepatic mitochondrial oxidative and anaplerotic fluxes in humans using dynamic  $^{13}\text{C}$  magnetic resonance spectroscopy. *Nature Med.* **20**, 98–102 (2014).
95. Farese, R. V., Jr, Zechner, R., Newgard, C. B. & Walther, T. C. The problem of establishing relationships between hepatic steatosis and hepatic insulin resistance. *Cell Metab.* **15**, 570–573 (2012).
96. Cohen, J. C., Horton, J. D. & Hobbs, H. H. Human fatty liver disease: old questions and new insights. *Science* **332**, 1519–1523 (2011).
97. Sun, Z. & Lazar, M. A. Dissociating fatty liver and diabetes. *Trends Endocrinol. Metab.* **24**, 4–12 (2013).
98. Niebergall, L. J., Jacobs, R. L., Chaba, T. & Vance, D. E. Phosphatidylcholine protects against steatosis in mice but not non-alcoholic steatohepatitis. *Biochim. Biophys. Acta* **1811**, 1177–1185 (2011).
99. Jacobs, R. L. *et al.* Impaired *de novo* choline synthesis explains why phosphatidylethanolamine N-methyltransferase-deficient mice are protected from diet-induced obesity. *J. Biol. Chem.* **285**, 22403–22413 (2010).
100. Xu, E. *et al.* Hepatocyte-specific Ptpn6 deletion promotes hepatic lipid accretion, but reduces NAFLD in diet-induced obesity: potential role of PPAR $\gamma$ . *Hepatology* **59**, 1803–1815 (2013).
101. Ruiz, R. *et al.* Sterol regulatory element binding protein-1 (SREBP-1) is required to regulate glycogen synthesis and gluconeogenic gene expression in mouse liver. *J. Biol. Chem.* **289**, 5510–5517 (2014).
102. Benhamed, F. *et al.* The lipogenic transcription factor ChREBP dissociates hepatic steatosis from insulin resistance in mice and humans. *J. Clin. Invest.* **122**, 2176–2194 (2012).
103. Brown, J. M. *et al.* CGI-58 knockdown in mice causes hepatic steatosis but prevents diet-induced obesity and glucose intolerance. *J. Lipid Res.* **51**, 3306–3315 (2010).
104. Cantley, J. L. *et al.* CGI-58 knockdown sequesters diacylglycerols in lipid droplets/ER-preventing diacylglycerol-mediated hepatic insulin resistance. *Proc. Natl Acad. Sci. USA* **110**, 1869–1874 (2013).
- The authors of this study demonstrate the importance of compartmentation of DAG in modulating hepatic insulin resistance.**
105. Bergman, B. C., Hunerdosse, D. M., Kerege, A., Playdon, M. C. & Perreault, L. Localisation and composition of skeletal muscle diacylglycerol predicts insulin resistance in humans. *Diabetologia* **55**, 1140–1150 (2012).
106. Sun, Z. *et al.* Hepatic Hdc3 promotes gluconeogenesis by repressing lipid synthesis and sequestration. *Nature Med.* **18**, 934–942 (2012).
107. Kumashiro, N. *et al.* Role of patatin-like phospholipase domain-containing 3 on lipid-induced hepatic steatosis and insulin resistance in rats. *Hepatology* **57**, 1763–1772 (2013).
108. Kumari, M. *et al.* Adiponutrin functions as a nutritionally regulated lysophosphatidic acid acyltransferase. *Cell Metab.* **15**, 691–702 (2012).
109. Speliotes, E. K. *et al.* Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet.* **7**, e1001324 (2011).
110. Lin, Y.-C., Chang, P.-F., Chang, M.-H. & Ni, Y.-H. Genetic variants in GCKR and PNPLA3 confer susceptibility to nonalcoholic fatty liver disease in obese individuals. *Am. J. Clin. Nutr.* **99**, 869–874 (2014).
111. Petersen, K. F. *et al.* Reversal of nonalcoholic hepatic steatosis, hepatic insulin resistance, and hyperglycemia by moderate weight reduction in patients with type 2 diabetes. *Diabetes* **54**, 603–608 (2005).
- This study demonstrates that moderate (<10%) body weight reduction in obese patients with type 2 diabetes eating a 1,200 calorie hypocaloric diet corrected fasting plasma glucose concentrations, normalized rates of hepatic glucose reversed NAFLD and reversed hepatic insulin resistance independent of changes in circulating adipocytokines.**
112. Lim, E. L. *et al.* Reversal of type 2 diabetes: normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol. *Diabetologia* **54**, 2506–2514 (2011).
113. Weiss, E. C., Galuska, D. A., Kettel Khan, L., Gillespie, C. & Serdula, M. K. Weight regain in U.S. adults who experienced substantial weight loss, 1999–2002. *Am. J. Prev. Med.* **33**, 34–40 (2007).
114. McGuire, M. T., Wing, R. R. & Hill, J. O. The prevalence of weight loss maintenance among American adults. *Int. J. Obes. Relat. Metab. Disord.* **23**, 1314–1319 (1999).
115. Haufe, S. *et al.* Long-lasting improvements in liver fat and metabolism despite body weight regain after dietary weight loss. *Diabetes Care* **36**, 3786–3792 (2013).
116. Mayerson, A. B., Hundal, R. S., Dufour, S. & Lebon, V. The effects of rosiglitazone on insulin sensitivity, lipolysis, and hepatic and skeletal muscle triglyceride content in patients with type 2 diabetes. *Diabetes* **51**, 797–802 (2002).
- This study demonstrated that thiazolidinediones improve insulin sensitivity in patients with type 2 diabetes by decreasing hepatic steatosis and promoting a redistribution of fat to the subcutaneous fat compartment.**
117. Kim, J. K. *et al.* Differential effects of rosiglitazone on skeletal muscle and liver insulin resistance in A-ZIP/F-1 fatless mice. *Diabetes* **52**, 1311–1318 (2003).
118. Prieur, X. *et al.* Thiazolidinediones partially reverse the metabolic disturbances observed in Bcl2/seipin-deficient mice. *Diabetologia* **56**, 1813–1825 (2013).
119. Dutchak, P. A. *et al.* Fibroblast growth factor-21 regulates PPAR $\gamma$  activity and the antidiabetic actions of thiazolidinediones. *Cell* **148**, 556–567 (2012).
120. Miyazaki, Y. *et al.* Rosiglitazone improves downstream insulin receptor signaling in type 2 diabetic patients. *Diabetes* **52**, 1943–1950 (2003).
121. Perry, R. J. *et al.* Reversal of hypertriglyceridemia, fatty liver disease, and insulin resistance by a liver-targeted mitochondrial uncoupler. *Cell Metab.* **18**, 740–748 (2013).
- The authors of this article demonstrate that a liver-targeted mitochondrial uncoupling agent (DNP) resulted in around a 60% increase in hepatic fat oxidation, reductions in liver and muscle triglyceride and diacylglycerol content and reversal of liver and muscle insulin resistance in rat models of NAFLD and type 2 diabetes.**

**Acknowledgements** Space limitations preclude this from being a comprehensive review, and this unfortunately limits appropriate recognition of many of our colleagues worldwide who have contributed immeasurably to the development of this field. This work was supported by grants from the National Institutes of Health: R01 DK-40936, R01 DK-49230, R24 DK-085836, R01 AG-23686, U24 DK-059635, UL1 RR-024139, P30 DK-45735, T32-DK101019 and I01-BX000901 (V.T.S.), and the Novo Nordisk Foundation for Basic Metabolic Research, University of Copenhagen.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at [go.nature.com/cghfru](http://go.nature.com/cghfru). Correspondence should be addressed to G.I.S. ([gerald.shulman@yale.edu](mailto:gerald.shulman@yale.edu)).

# Pro-resolving lipid mediators are leads for resolution physiology

Charles N. Serhan<sup>1</sup>

**Advances in our understanding of the mechanisms that bring about the resolution of acute inflammation have uncovered a new genus of pro-resolving lipid mediators that include the lipoxin, resolvins, protectins and maresins families, collectively called specialized pro-resolving mediators. Synthetic versions of these mediators have potent bioactions when administered *in vivo*. In animal experiments, the mediators evoke anti-inflammatory and novel pro-resolving mechanisms, and enhance microbial clearance. Although they have been identified in inflammation resolution, specialized pro-resolving mediators are conserved structures that also function in host defence, pain, organ protection and tissue remodelling. This Review covers the mechanisms of specialized pro-resolving mediators and omega-3 essential fatty acid pathways that could help us to understand their physiological functions.**

Excessive inflammation is widely appreciated to be a unifying component in many chronic diseases, including vascular diseases, metabolic syndrome and neurological diseases, and thus is a public health concern. Understanding endogenous control points within the inflammatory response could provide us with new perspectives on disease pathogenesis and treatment approaches. Barrier break, trauma and microbial invasion create a need for the host to neutralize invaders, clear the site, and remodel and regenerate tissue. The acute inflammatory response is protective, providing a terrain in which lipid mediators — such as eicosanoids (prostaglandins and leukotrienes)<sup>1,2</sup> produced from the essential fatty acid arachidonic acid, as well as many cytokines and chemokines<sup>3–5</sup> — have crucial roles in the initial response. Interactions among prostaglandins, leukotrienes and pro-inflammatory cytokines amplify inflammation, the signs of which can be reduced by pharmacological inhibition and receptor antagonists<sup>1–3</sup>. However, given that excessive inflammation contributes to many widely occurring diseases, improvements are needed in treatment and in our understanding of the mechanisms involved.

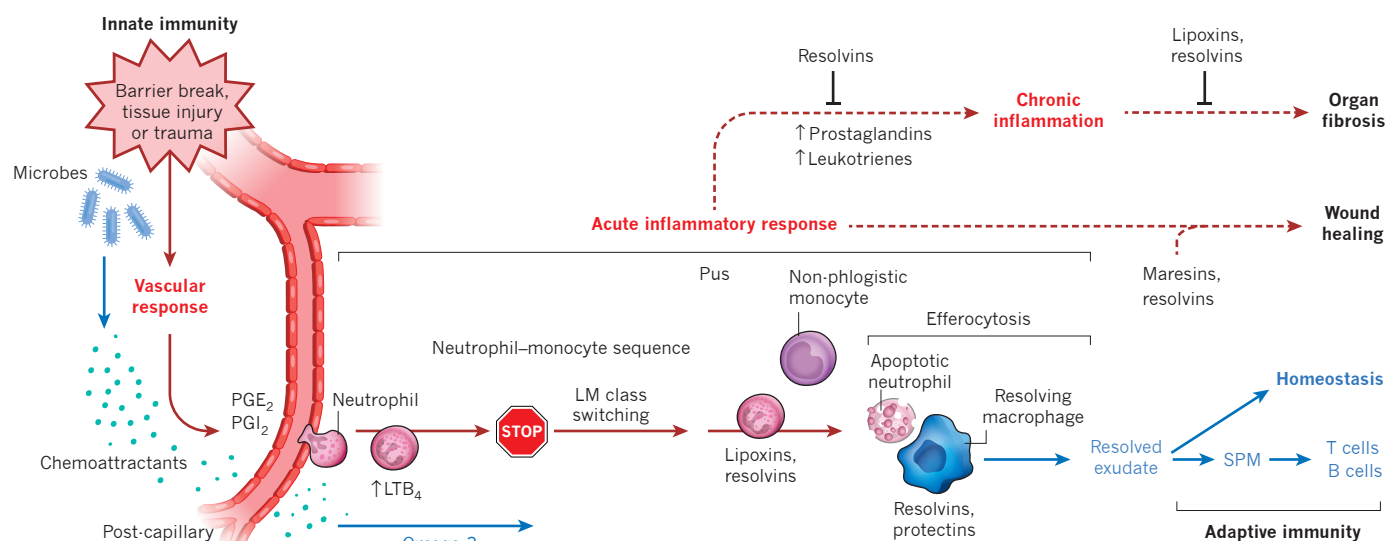
Pathologists divide the acute inflammatory response into initiation and resolution (Fig. 1). Resolution was considered to be a passive process<sup>6</sup>. With the identification of mediators with pro-resolving capacity that could be biosynthesized from omega-3 (or n-3) essential fatty acids (EFA), evidence emerged that resolution of self-limited acute inflammation might be an active, programmed response that is ‘turned on’ in animal models, and not simply a process of passive dilution of chemoattractants<sup>7,8</sup>. For a metabolite to fulfil the role of mediator, it must be produced in sufficient amounts *in vivo* to evoke bioactions. The omega-3 fatty acids EPA and DHA, which are found in marine oils, have long been thought to have anti-inflammatory properties, whereby they compete with arachidonic acid, reducing pro-inflammatory eicosanoids<sup>9</sup>. The molecular mechanism by which this occurs is unclear, and the evidence is inconclusive as to whether omega-3 EPA and DHA are beneficial for human health and as treatments for disease. Resolving inflammatory exudates use omega-3 fatty acids (Fig. 1) to produce structurally distinct families of signalling molecules — resolvins, protectins and maresins, collectively termed specialized pro-resolving mediators

(SPMs). This realization prompted new interest in resolution pathways and the innate immune mechanisms for homeostasis. SPMs are agonists with the potential to stimulate key cellular resolution events, namely limiting polymorphonuclear neutrophil infiltration and enhancing macrophage clearance of apoptotic cells<sup>5,6</sup>, as shown in pre-clinical studies using animal models<sup>10</sup>. This Review addresses the role of novel lipid-derived SPMs in resolution that also function in host defence, pain and tissue regeneration, which could help us to understand the role of SPM pathways in human resolution physiology.

## Cellular events in resolution of acute inflammation

During the initiation phase, leukocytes traffic from circulation to the site of trauma or microbial invasion, forming inflammatory exudates — conventionally these are considered to be the ‘battlefields’ where most resolution occurs<sup>6</sup>. The first responders, neutrophils, swarm like bees to defend the body, moving along chemotactic gradients (for example, increasing levels of the leukotriene LTB<sub>4</sub>)<sup>11</sup>, and exiting venules governed by prostaglandins (such as PGE<sub>2</sub> and PGI<sub>2</sub>) that act on vascular cells and blood flow<sup>1</sup> (Fig. 1). These lipid mediators, along with many cytokines, chemokines and complement components (C5a and C3b), stimulate the chemotaxis of neutrophils into tissues to phagocytize and neutralize invaders<sup>3–6</sup>. Many therapeutic agents block or antagonize the initiation steps of acute inflammation (for example, prostaglandin biosynthesis inhibitors or chemokine receptor antagonists)<sup>1,3,12</sup>. At a cellular level, the main events of resolution are the cessation of neutrophil influx and efferocytosis (macrophage clearance of debris, including apoptotic neutrophils)<sup>4,5</sup>. The protective acute inflammatory response evolved to permit the repair of injured tissues and eliminate invading organisms. Ideally, it is a self-limited process, leading to complete resolution that enables a return to homeostasis (Fig. 1). By studying self-limited inflammation in animal models, and using a systems approach to investigate resolving exudates, novel bioactive products derived from essential fatty acids were uncovered. The bioactions of these products include limited neutrophil influx *in vivo*, reduced human neutrophil transmigration and counter-regulation of cytokines such as tumour necrosis factor- $\alpha$  (TNF- $\alpha$ ) in mice. Each bioactive product

<sup>1</sup>Center for Experimental Therapeutics and Reperfusion Injury, Department of Anesthesiology, Perioperative and Pain Medicine, Harvard Institutes of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA.



**Figure 1 | Lipid mediators in the acute inflammatory response, resolution and other outcomes.** Lipid mediators have pivotal roles in the vascular response and leukocyte trafficking, from initiation to resolution. Lipoxins, resolvins, protectins and maresins, collectively called specialized pro-resolving mediators (SPMs), are produced during this self-limited response (Fig. 2). After barrier break, tissue injury or trauma, eicosanoids are crucial in initiating the cardinal signs of inflammation (redness, heat, pain and swelling). As part of the vascular response, leukocytes traffic to the site of injury. The prostaglandins  $\text{PGE}_2$  and  $\text{PGI}_2$  (involved in vasodilation) and the leukotriene  $\text{LTB}_4$  (involved in chemotaxis and adhesion) stimulate the migration of neutrophils to the tissue. As part of the neutrophil–monocyte sequence, lipid mediator (LM) class switching from eicosanoids to lipoxins signals ‘stop’ and begins the end of the acute inflammatory response. Lipoxins and resolvins stimulate the recruitment of non-phlogistic monocytes. The resolving macrophages then clear apoptotic

neutrophils in a process called efferocytosis (stimulated by resolvins and protectins). After this has taken place normal structure and homeostasis can be restored. Signs of resolution include sequestration of pro-inflammatory cytokines, clearance of neutrophils from epithelial surfaces, phagocytosis of apoptotic neutrophils and removal of inflammatory debris and microbial invaders. Failed resolution can lead to increased levels of prostaglandins and leukotrienes, chronic inflammation (which can be inhibited by resolvins) and fibrosis. SPMs counter-regulate pro-inflammatory chemical mediators, reducing the magnitude and duration of inflammation, and stimulate re-epithelialization, wound healing and tissue regeneration in model organisms. In addition to the release of omega-3 substrates from phospholipid stores<sup>95</sup>, these substrates can enter exudate as a result of oedema from peripheral blood, as shown in mice<sup>57</sup>. SPMs enhance efferocytosis, stimulate signs of resolution and signal to adaptive immunity.

was systematically evaluated, and found to be biosynthesized from EPA and DHA by separate pathways in human neutrophils and macrophages *in vitro*<sup>7,8,13</sup>.

### Elucidation of pro-resolving lipid-derived mediators

The anti-inflammatory process is not equivalent to pro-resolution, which involves SPMs acting as agonists to stop further neutrophil influx and the activation of nonphlogistic responses by macrophages and resolution programs (Fig. 1). The key to understanding this difference was the identification of novel families of autacoids and forms of these molecules that could potentially be triggered by aspirin, providing evidence that, in animal models, resolution is actively orchestrated by lipid mediators (Fig. 2 and Box 1). Harnessing the pathways that stimulate resolution will be a challenge for the future<sup>14</sup>.

Omega-3 fatty acids are a widely taken dietary supplement, but less than 25% of these supplements are prescribed by health-care providers<sup>15</sup>. Given the mixed results<sup>16</sup> on the benefit of these supplements in clinical trials, it is crucial for public health that the mechanisms that underlie their requirement are established.

Using a systems approach with resolving mouse exudates was key to elucidating SPM actions and pathways<sup>4,8</sup>. Biosynthesis with human leukocytes and complete stereochemistry of each major resolvins (RvE1, RvD1, RvD2, RvD3 and RvD5), protectin and maresin was accomplished by matching these mediators with those prepared by total organic synthesis, which also confirmed their potent bioactivity<sup>10,17</sup>. Resolvins and protectins control the magnitude and duration of inflammation in animal disease models<sup>18</sup>, and can, for example, increase animal survival<sup>10</sup> (Fig. 2). The potent *in vivo* actions of RvD1 and RvD2 are reported in many pathologies, such as obesity and pathologies affecting the vascular<sup>19</sup>, airway<sup>20</sup>, dermal, renal and ocular systems, and in processes including pain, fibrosis and wound healing<sup>10,21</sup>. Their role in governing neutrophil influx, resolution

macrophages and reducing pro-inflammatory mediators (PAF,  $\text{LTB}_4$  and prostaglandins) seem to be fundamental in all organs.

### Resolvin D3 and the maresin pathway

Within self-limited exudates, RvD3 has a unique time frame compared with RvD1 and RvD2. In mouse peritonitis, RvD3 appears late in resolution, suggesting it has a specific role. The complete stereochemistry of RvD3 was recently established<sup>22</sup>, confirming its potent anti-inflammatory and pro-resolving actions<sup>8</sup>. Macrophage biosynthesis of maresin MaR1 and its potent pro-resolving and tissue regenerative actions<sup>23</sup> (Fig. 2) involve an active 13S,14S-epoxide-maresin intermediate, stimulating a macrophage phenotype switch from M1 to M2 (ref. 24) (Fig. 2). The switch towards the M2 phenotype is associated with reparative and anti-inflammatory macrophage functions<sup>5,25</sup>.

### Resolution agonists and resolution disrupters

Several lessons have emerged from recent studies. Prostaglandins are central to vascular responses, permitting neutrophils and monocytes to leave post-capillary venules (a process called diapedesis). Their production through COX-1 and COX-2 (also known as prostaglandin G/H synthase 1 and 2) is crucial for initiation and timely resolution<sup>26,27</sup> (Fig. 1).  $\text{PGE}_2$  and  $\text{PGD}_2$  each evoke pro-inflammatory and anti-inflammatory responses that depend on tissue location<sup>12</sup>.  $\text{PGE}_2$  enhances  $\text{LTB}_4$ -mediated neutrophil extravasation and tissue injury that is blocked, for example, by topical administration of synthetic lipoxin  $\text{A}_4$  ( $\text{LXA}_4$ ) and its aspirin-triggered epimer 15-epi- $\text{LXA}_4$  (ref. 28), illustrating both a pro-inflammatory  $\text{PGE}_2$  function in mouse skin and the ability of 15-epi- $\text{LXA}_4$  mimetics to limit neutrophil infiltration and tissue injury. Liquid chromatography with tandem mass spectrometry (LC–MS–MS)-based profiling demonstrated the temporal switch from an environment with high levels



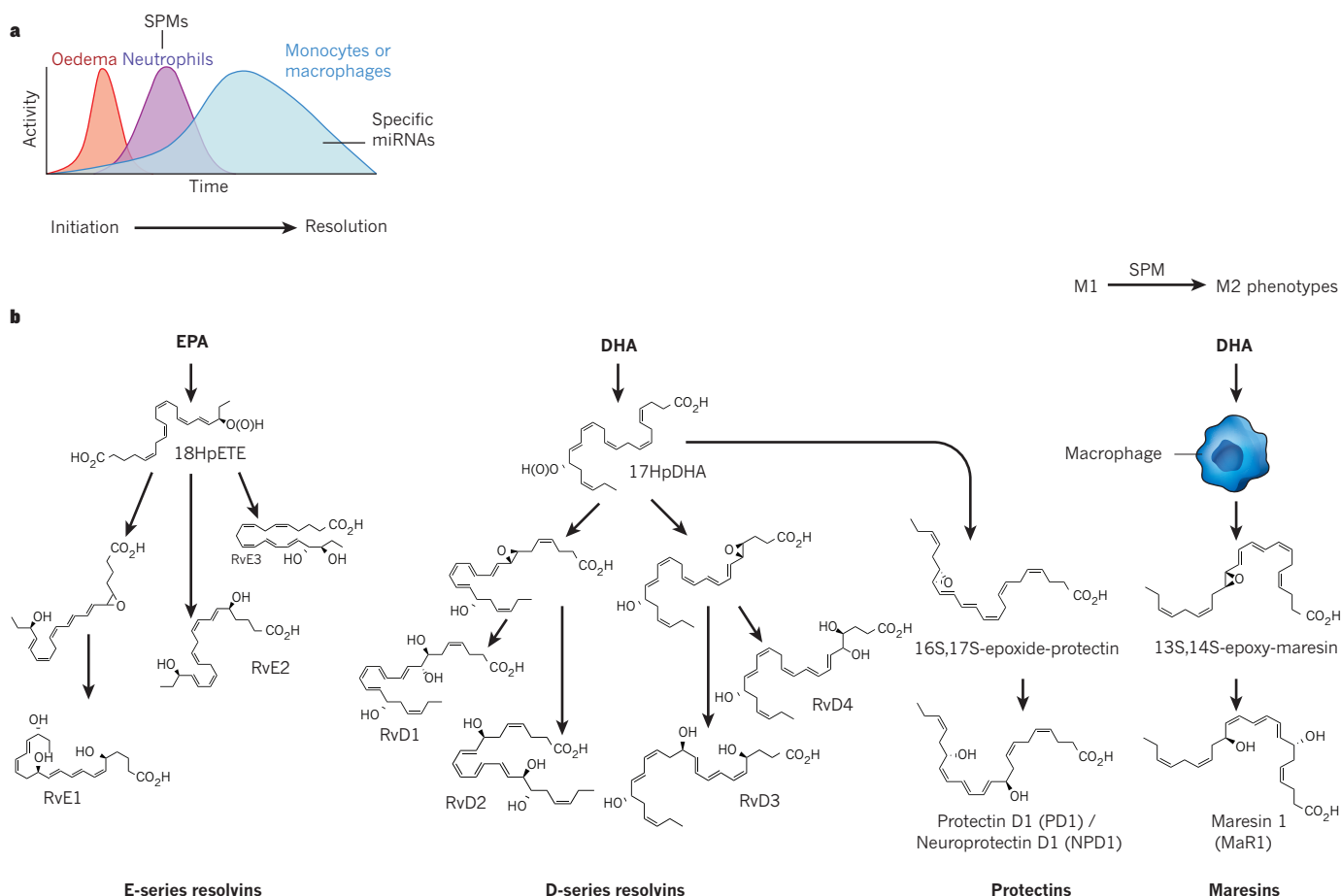
of prostaglandins and  $LTB_4$  to one with high levels of lipoxins, a process known as lipid-mediator class switching, in mouse exudates (Fig. 1 and Box 1).  $PGE_2$  or  $PGD_2$  added to isolated human neutrophils increases 15-lipoxygenase type I translation from messenger RNA stores in a cyclic-AMP-dependent manner, increasing lipoxin biosynthesis (identified using MS–MS spectra)<sup>26</sup>.

Inhibition of COX-2 delays resolution because prostaglandins have crucial roles in resolution and because they are initiators of lipid-mediator class switching (Fig. 1), as shown in animal disease models *in vivo*<sup>18,26,29</sup>. When mapping resolution, it became apparent that initiation signals the end of inflammation<sup>4</sup> and that leukocyte traffic in pus permits prostanoids to signal the biosynthesis of other resolution mediators (Fig. 1 and Fig. 2). For example, disruption of physiological lipid-mediator class switching has deleterious consequences in mouse models of arthritis<sup>29</sup>.

To pinpoint the crucial steps and mechanisms of SPM action within inflammation resolution, it was important to introduce quantitative indices<sup>18,30</sup> that enabled the assessment of resolution *in vivo*<sup>21,27,31</sup>. Resolution indices identified agents that stimulate as well as those that disrupt or delay resolution (resolution interval); for example, COX-2 and lipoxygenase inhibitors<sup>18,27,32</sup>. Specific SPMs shorten the resolution interval by limiting neutrophil recruitment and stimulating both macrophage efferocytosis (Fig. 1) and bacterial killing<sup>31,33,34</sup>, demonstrating the neutrophil–monocyte sequence and the macrophage responses needed for tissue regeneration<sup>23</sup>.

Glucocorticoids, specific cyclin-dependent kinase inhibitors, statins, annexin peptides and aspirin all enable resolution<sup>31,35,36</sup>. In the same way there are many mediators in the initiation of inflammation, there are also many endogenous mediators and drugs that have an impact on resolution<sup>5,18,20</sup>.

Although both aspirin and non-steroidal anti-inflammatory drugs (NSAIDs) inhibit prostanoid biosynthesis, aspirin is an irreversible inhibitor that acetylates COX-1 and COX-2, whereas other NSAIDs are reversible inhibitors<sup>1,2</sup>. Aspirin acetylation of COX-2 modifies the catalytic domain, blocking prostaglandin biosynthesis<sup>1,2</sup>, but it remains active, producing 15R-HETE from arachidonic acid, 18R-HEPE from EPA and 17R-HDHA from DHA in cells carrying COX-2. These products can be transformed by human neutrophils *in vitro* to aspirin-triggered lipoxins, aspirin-triggered resolvins<sup>7,8</sup> and aspirin-triggered protectins<sup>37</sup>. Each has a potent effect, stopping human neutrophil migration and enhancing macrophage clean up, improving resolution in mice. Whether aspirin or statins enhance the production of aspirin-triggered SPMs in humans remains to be established using mass spectral-based identification. And whether aspirin favours resolution in humans, in whom distinct resolution-phenotypes have emerged, is of considerable interest<sup>27,38</sup>. In mice, intravascular  $LXA_4$  is produced by platelet–neutrophil aggregates during ischaemia, which reduces vascular inflammation. Aspirin triggers 15-epi- $LXA_4$ , which was identified using LC–MS–MS, and this was less effective in mice deficient for the lipoxin receptor ALX,



**Figure 2 | Production of specialized pro-resolving mediators in resolving inflammatory exudates.** **a**, A typical self-limited acute inflammatory response time course in experimental settings from initiation (time 0) to resolution: oedema; neutrophilic infiltration, including the temporal biosynthesis of specialized pro-resolving mediators (SPMs) that occurs in resolving exudates; and non-phlogistic recruitment of monocytes or macrophages, which is required for homeostasis, and the repair and

regeneration of injured tissues (involving specific microRNAs regulated by SPM receptors). **b**, Resolution omega-3 metabolome. Biosynthesis of resolvins, protectins and maresins from EPA and DHA with the main bioactive structures from each family (Box 1; see refs 17 and 20 for details of the biosynthetic mechanisms and stereochemical assignments of the bioactive products). Each SPM stimulates macrophage switching to the M2 phenotype and is produced by human neutrophils, apoptotic neutrophils and macrophages<sup>13</sup>.

## BOX 1

## Resolution omega-3 metabolome

Bioactive products of omega-3 were originally isolated from mouse resolving exudates. The structure of these products was elucidated, and the biosynthesis of each new omega-3 family from EPA and DHA was recapitulated with isolated human neutrophils<sup>8,20</sup> and macrophages *in vitro*<sup>23,24</sup>. Biosynthesis of E-series resolvins is initiated by the molecular oxygen insertion at the carbon-18 position of EPA produced by hypoxic human endothelial cells and acetylated COX-2, or p450. Microbial p450 (ref. 7) (Fig. 2, left) also produces 18-HEPE, which is converted to bioactive E-series members by human neutrophils. Both aspirin-dependent and independent formation occurs in human peripheral blood<sup>47,61</sup>. The resolution metabolome also activates 17-lipoxygenation of DHA by hydrogen abstraction; 17S-HpDHA is converted to resolvins-epoxide-containing intermediates (Fig. 2, centre) that are transformed to resolvins D1–D4 by human neutrophils, each identified using methods such as mass spectrometry of bioactive products. The 17S-HpDHA intermediate is also a precursor to 16,17-epoxide-protectin, which is converted

to protectin D1/neuroprotectin D1 and related structures by human leukocytes (neutrophils and T cells), neural cells and retinal-pigmented epithelial cells<sup>8,97,98,112</sup>. Maresins isolated from human and mouse macrophages are produced by initial lipoxygenation with molecular oxygen insertion at the carbon-14 position to form the hydroperoxide intermediate, which is rapidly converted to 13S,14S-epoxide-maresin and enzymatically converted to maresin-1 (ref. 24) (Fig. 2, right). Lipoxygenase mechanisms involve hydrogen abstraction and molecular oxygen insertion at specific carbon positions that are predominantly in the S configuration. Aspirin, by COX-2 acetylation and p450 enzymes, contributes to the biosynthesis of R-configuration alcohols in lipoxins, resolvins and protectins<sup>37</sup>. The stereochemistry of each bioactive specialized pro-resolving mediator family member (Fig. 2) has been determined, and the biosynthesis and potent pro-resolving and anti-inflammatory actions in murine exudates and human tissues has been confirmed (see ref. 17 for original reports and refs 10, 20 for reviews).

providing *in vivo* evidence that aspirin can jump-start resolution circuits in mice<sup>39</sup>.

## New mechanisms in local SPM biosynthesis

Microparticles are membrane-derived vesicles produced by a range of cell types that contribute to human pathologies. Microparticles from self-resolving exudates show anti-inflammatory and pro-resolving capacity<sup>40</sup> in mice. Resolution microparticles enhance efferocytosis<sup>13,40</sup> and carry pro-resolving signals, including hydroxy-SPM-intermediates esterified in phospholipids<sup>40</sup>. Secreted PLA<sub>2</sub> (sPLA<sub>2</sub>) prompts the release of pro-resolving signals from microparticles for trans-cellular conversion by human macrophages *in vitro*<sup>13,40</sup>. Because nanomedicines are useful for local targeting and delivery, resolution microparticles, and their ability to shorten the resolution interval in mouse peritonitis, were used as a basis for biomimicry to construct humanized nanoparticles containing an LXA<sub>4</sub> analogue, or aspirin-triggered-RvD1 (ref. 40). These nano-pro-resolving medicines (NPRMs), carrying SPMs or SPM analogues, enhance wound healing of human keratinocytes and are protective in a mouse model of temporomandibular joint disease characterized by inflammation-induced bone loss<sup>40</sup> (Fig. 1 and Box 2).

Microparticles can also transfer substrate and intermediates to macrophages during efferocytosis to enhance SPM biosynthesis, which was demonstrated by transfer of deuterium labels from precursors to SPMs (identified using LC–MS–MS)<sup>13</sup>. Myeloid cells at different stages have agonist- and phenotype-specific lipid mediator profiles. For example, human neutrophils from healthy peripheral blood predominantly produce LTB<sub>4</sub>, whereas apoptotic neutrophils produce PGE<sub>2</sub>, LXB<sub>4</sub> and RvE2 signals<sup>13</sup>.

Both M1 and M2 macrophages have specific markers and pathways that are specialized to the functions of that subpopulation in inflammation and resolution<sup>25</sup>. Human M2 macrophages have more of the enzymes<sup>41</sup> needed for cell-type-specific lipid mediators. M2 cells produce SPM profiles with lower levels of LTB<sub>4</sub> and prostaglandins than M1 cells. Both cell types engulf apoptotic neutrophils, changing their lipid mediator signature profiles. In M2 cells, LTB<sub>4</sub> is downregulated and SPMs are increased<sup>13</sup>, suggesting that M1 and M2 subpopulations<sup>25,41</sup> produce functional lipid mediator signatures that can affect both physiological and pathophysiological states<sup>13</sup>. In addition, group IID sPLA<sub>2</sub> was identified as a resolving sPLA<sub>2</sub> that was expressed in dendritic cells and macrophages that release substrates with the capacity for producing RvD1 and PGJ<sub>2</sub>, as identified by mass spectrometry in mouse lymphoid

tissue *in vivo*<sup>42</sup>. The functional contributions of this resolving sPLA<sub>2</sub> to the inflammatory response in humans is unknown.

The role of eosinophils is well appreciated in parasitic infections and allergic responses. In patients with severe asthma, protectin D1 (PD1) is present in exhaled breath condensates<sup>43</sup>, and is lower in eosinophils from these patients<sup>44</sup>. Human eosinophils produce PD1, which reduces adhesion molecules (CD11b and L-selectin), eotaxin-1 and chemotaxis at nanomolar concentrations, without affecting degranulation, superoxide generation or cell survival. Eosinophils also stimulate resolution in mouse peritonitis through SPMs initiated by mouse eosinophils<sup>45</sup>. LC–MS–MS-lipidomics identified LXA<sub>4</sub>, RvD5, 17-HDHA and PD1 from eosinophils, and RvE3 *in vivo* (Fig. 2), which limit neutrophil infiltration and regulate macrophages<sup>44–46</sup>. Hence, by their ability to produce SPMs, eosinophils might contribute to resolution. In support of this, eosinophil depletion has been found to lead to failed resolution, which is rescued by PD1 or eosinophil restoration in mice. Thus, cellular traffic to sites of inflammation has a dynamic impact on lipid mediator signatures and specific SPM metabolomes activated within the local milieu.

## SPM cellular actions in disease models

SPMs increase survival in diverse mouse models; for example, inflammation and tissue injury in the airway, skin and eye that result from collateral damage are improved with exogenous SPMs<sup>10</sup>. The SPM nanomolar doses required to stop ongoing inflammation and promote resolution rely on G-protein-coupled receptors (GPCRs). Several SPM receptors have been identified using GPCR screening, labelled ligands for specific binding (stereospecific nanomolar dissociation constant) and functional cellular responses<sup>47–49</sup>. In general, SPMs do not utilize intracellular calcium (Ca<sup>2+</sup>) mobilization in neutrophils for signal transduction but instead activate phosphorylation. RvE1 specifically binds to the receptors ChemR23 (ref. 47) and BLT1 to evoke pro-resolving responses. RvE1 activation of ChemR23 enhances macrophage phagocytosis by phosphoprotein-mediated signalling<sup>48</sup>. RvE1 blocks LTB<sub>4</sub> binding and promotes, through BLT1, apoptosis of neutrophils for their clearance by macrophages<sup>49</sup>, whereas LTB<sub>4</sub>–BLT1 signals neutrophil survival. The blocking of survival signals to neutrophils by RvE1 marks an important difference from the process that occurs in the innate response, in which neutrophils must undergo timely apoptosis for clearance<sup>5,49</sup> (Fig. 2).

RvD1 binds and activates the human receptor GPR32, and binds to both the human and murine forms of the LXA<sub>4</sub> receptor (called ALX/FPR2). Transgenic mice overexpressing human ALX/FPR2 require

## BOX 2

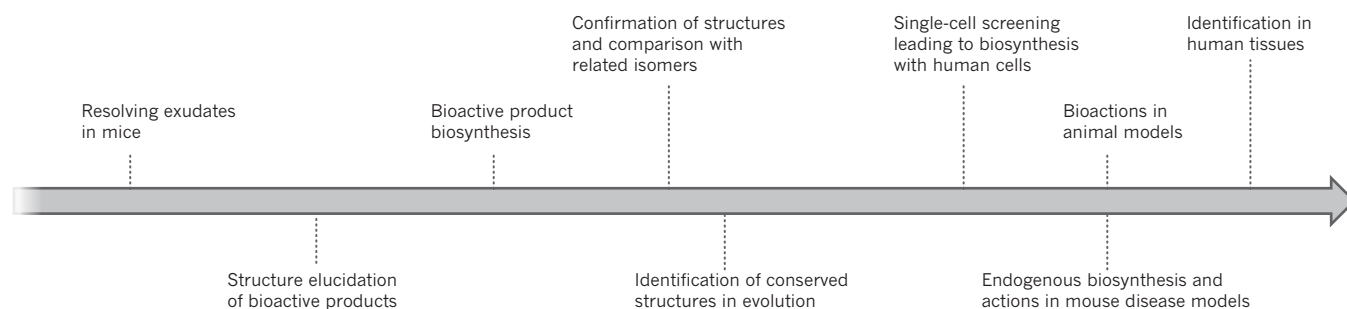
# Omega-3 pro-resolving mediator structure, function and steps towards human translation

A number of steps have led to the possibility of assessing SPM function in humans (Box 2 Fig.). The molecules derived from omega-3 and their bioactions have been identified in the resolving inflammatory exudates of mice. Studies have reported the structure of these molecules, their actions with human neutrophils and their proposed biosynthesis from EPA and DHA using 18-oxygen, 14-carbon and trapping of transient intermediates<sup>17</sup>. Confirmation of the bioactive structures and total organic synthesis to assign complete stereochemistry to each specialized pro-resolving mediator (SPM) family member required the development of a strategy to match the biologically active isolated product with those prepared by total organic synthesis. NMR (nuclear magnetic resonance) confirmed the geometry of the conjugated double bond systems present in SPMs; for example, RvE1 (ref. 47), RvD1, RvD2 (ref. 34) or MaR1 (refs 23, 24) (ref. 17 describes the matching criteria). The resolving functions of SPMs are separate from those of anti-inflammatories; to serve as an SPM, the omega-3 product must be biosynthesized at the *in vivo* levels that are commensurate with the SPM's bioactions. For many of the SPMs, this criterion has been achieved and confirmed with commercial resolvins (see ref. 10 for a review).

SPMs have been identified as conserved structures that are present in trout (PD1 and RvD1)<sup>88</sup>, salmon (RvD1 and RvD2)<sup>117</sup> and planaria (RvE1 and MaR1)<sup>23</sup>. Microfluidic chambers that permit visualization of cell–cell interactions between leukocyte subpopulations (human neutrophils and monocytes) and distinguish phlogistic compared with non-phlogistic phagocyte behaviour are ideal for screening SPMs and humanized nano-pro-resolving medicines (NPRMs)<sup>40,56</sup>. This single-cell

screening permits the optimization of enrichment of microparticles with SPMs, the production of NPRMs and the viewing of neutrophil–monocyte interactions<sup>56,57</sup>, which are essential for appreciating the signals that occur during the neutrophil–monocyte sequence (Fig. 1).

Using microfluidic chambers, single human neutrophils were assayed within about 5 minutes of their capture from whole blood (minimizing the isolation time reduces potential artefacts). Nanomolar concentrations of each SPM prevents human neutrophil migration to interleukin-8 (IL-8), whereas at equimolar doses the precursor (DHA) is not active<sup>56,57</sup>. Resolution indices permit the quantitative assessment of SPM actions in animal models<sup>18,30</sup>; this is essential for defining the SPMs within the integrated response of the host to acute inflammatory challenges. Demonstration and identification of omega-3 SPMs in human tissues is required so that we can appreciate their potential roles in humans. RvE1 and RvE2 (refs 47, 61) were identified in the peripheral blood of healthy volunteers, some of whom had been given EPA supplements, using LC–MS–MS multi-reaction monitoring. RvD1, RvD2 (ref. 111), PD1 and 17-HDHA were identified in the exhaled breath condensates of people with asthma<sup>43</sup>; and additional SPMs were identified in human adipose tissue<sup>118</sup>, the brains of people with Alzheimer's disease<sup>96,101</sup>, and patients with multiple sclerosis<sup>119</sup> and rheumatoid arthritis<sup>120</sup> using mass spectral identification. Although, at present, the demonstration of omega-3 in human tissues is in the initial stages, with the capabilities of LC–MS–MS-based profiling<sup>13</sup>, SPMs could have potential as markers for nutritional status. The field is now set with tools for assessing SPM function in humans and the relationships of SPMs with nutrition and disease.



less RvD1 than wild-type mice to stop inflammation<sup>50</sup>, and in receptor-deficient mice, RvD1 seems to have no leukocyte-directed actions<sup>51</sup>. Resolution involves specific microRNAs (miRNAs) regulated by SPM receptors<sup>50,52,53</sup>. RvD1–GPR32 upregulates miR-208 and anti-inflammatory interleukin-10 (IL-10), and downregulates miR-219, decreasing LTB<sub>4</sub> levels through the regulation of 5-lipoxygenase<sup>52</sup>. This miRNA regulation is an example of SPM signalling with a sustained tissue impact.

SPM receptors are also rapid signalers. For example, recombinant RvD1–GPR32 blocks histamine H1 receptor-stimulated increases in intracellular Ca<sup>2+</sup> in Chinese hamster ovary (CHO) cells by rapidly stimulating the phosphorylation of the H1 receptor, stopping Ca<sup>2+</sup> mobilization<sup>54</sup>. This form of SPM signalling, which has been reported in conjunctival goblet cells and RvD1, is also functional in salivary glands<sup>55</sup> and is probably relevant in human neutrophils, which rapidly stop chemotaxis in microfluidic chambers and change shape on exposure to SPMs<sup>56,57</sup>. RvD3 and RvD5 can also activate recombinant GPR32 (refs 22, 33). Given the temporal production of RvD3 *in vivo*<sup>22</sup> (Fig. 2), these findings

underscore that locally produced SPMs can affect different cell types and receptors with a spatio-temporal dependency.

In addition to RvD1 and LXA<sub>4</sub>, ALX/FPR2 is activated by peptide pro-resolving mediators, such as annexin A1, as well as high concentrations of pro-inflammatory peptides<sup>36</sup>. This capacity of ALX/FPR2 involves ligand-biased receptor activation and heterodimerization of ALX/FPR2, with related FPRs dictating pro-inflammatory signalling and homodimerization of ALX/FPR2 determining pro-resolving signalling<sup>36</sup>. LXA<sub>4</sub> also enhances the activity of the ALX/FPR2 promoter, which is mutated in human cardiovascular disease<sup>58</sup>.

## Infection and resolution programs

Topical RvE1 and LXA<sub>4</sub> each reduce the severity of periodontal disease in rabbits by enhancing clearance of the causative organism *Porphyromonas gingivalis*<sup>59,60</sup>. Although the anti-inflammatory actions of SPMs were uncovered in sterile inflammation models<sup>7,8</sup>, the relationship between resolution and infection is of interest because of the known eventual



immunosuppressive actions of anti-inflammatory drugs<sup>12</sup>. Surprisingly, RvD2 protects mice from caecal ligation puncture (CLP)-induced sepsis<sup>34</sup>; it has potent actions, enhancing phagocytosis and bacterial killing. In self-limited *Escherichia coli* infections, resolution programs are activated in mice, and levels of host PD1, RvD5 and RvD1 are elevated<sup>33</sup>. When SPMs identified *in vivo* were added back to mouse phagocytes, human macrophages or neutrophils as synthetic SPMs, they enhanced bacterial phagocytosis and killing, as well as clearance<sup>33,34,61</sup>. Importantly, when the host is given SPMs that are known to act on the host, lower antibiotic doses are needed to clear infections<sup>33</sup>.

Exogenous LXA<sub>4</sub> is also protective in rat CLP-induced sepsis, reducing bacterial burden and pro-inflammatory mediators through a macrophage nuclear-factor- $\kappa$ B (NF- $\kappa$ B)-mediated mechanism, which reduces systemic inflammation<sup>62</sup>. Aspirin-triggered-LXA<sub>4</sub> increases *E. coli* phagocytosis in a phosphatidylinositol-3-OH kinase (PI(3)K)- and scavenger receptor-dependent manner, and ALX/FPR2 is upregulated in patients with Crohn's disease and enhances bacterial clearance<sup>63</sup>. *Mycobacterium tuberculosis* infections are susceptible to modulation of LTA<sub>4</sub> hydrolase. Besides altering LTB<sub>4</sub>, this may also engage resolution programs by activating LTB<sub>4</sub>-LXA<sub>4</sub> production, regulating host responses in zebrafish, mice and possibly in humans<sup>64,65</sup>. Given the problem of rising antibiotic resistance, activation of resolution programs could provide new antimicrobial approaches to lower our exposure to antibiotics<sup>33</sup>.

The herpes simplex virus causes ocular infections that lead to stromal keratitis with viral-initiated immunopathology. RvE1 and PD1 are both potent and topically active in mouse models of the infection, reducing pro-inflammatory mediators and stimulating IL-10 (refs 66, 67). Lethal dissemination of the H5N1 virus activated genes in mice that are associated with lipoxin biosynthesis, whereby sustained inflammation inhibited lipoxin-mediated anti-inflammatory host responses, permitting viral dissemination<sup>68</sup>. H5N1 activates the host resolution-metabolome increasing PD1 levels (identified by LC-MS-MS)<sup>69</sup>. Host protectins show antiviral activity, blocking replication of the H5N1 influenza virus. During the time course of infection with the low-pathogenicity strain of influenza H3N2, anti-inflammatory mediators are produced that correlate with resolution and SPM-related pathway markers<sup>70</sup>. SPMs are also found in yeast infections; for example, *Candida albicans* infections, in which RvE1 enhances yeast killing and clearance in mice<sup>71</sup>. These results support the idea that treating the infected host with host-directed pro-resolving molecules could provide new opportunities to reduce inflammation and enhance clearance by exploiting host-pathogen interactions<sup>33,61,72</sup>.

### Chronic inflammatory disease models

Periodontal disease is a chronic inflammatory disease in which infection leads to neutrophil-mediated tissue injury around the tooth. Activated neutrophils from patients with periodontitis produce PGE<sub>2</sub>, LTB<sub>4</sub> and LXA<sub>4</sub>, each identified using LC-MS-MS<sup>73</sup>. PGE<sub>2</sub> in this tissue leads to bone loss. *P. gingivalis* elicits neutrophil recruitment in mouse air pouches along with COX-2 upregulation. LXA<sub>4</sub> stable analogues reduced both neutrophil influx and COX-2 expression that was associated with the oral pathogen<sup>73</sup>. In addition, *P. gingivalis* increased COX-2 expression in the lung and heart of mice, and 16S ribosomal RNA of *P. gingivalis* was present in these tissues, providing evidence of a role for this oral pathogen in the development of systemic inflammation. Transgenic rabbits overexpressing human 15-lipoxygenase type I produce 6–10 times the amounts of LXA<sub>4</sub>, which was identified using LC-MS-MS, than wild-type rabbits<sup>59</sup>; transgenic rabbits have less bone loss in periodontitis and markedly reduced neutrophil recruitment and vascular leakage through their skin on immune challenge compared with wild-type rabbits, suggesting that overexpression of lipoxin biosynthesis is protective and could be useful in controlling inflammation-mediated bone destruction<sup>59</sup>.

Unexpectedly, overexpression of 15-lipoxygenase in these transgenic rabbits sharply reduces atherosclerotic lesions<sup>74</sup>. In 12/15-lipoxygenase transgenic mice, LC-MS-MS identified RvD1, PD1 and 17-HDHA with reduced PGE<sub>2</sub> from activated macrophages<sup>75</sup>. LXA<sub>4</sub>, PD1 and RvD1 each reduced cytokines (for example, MCP-1) from endothelial cells

and adhesion molecules (P-selectin and VCAM-1) but not ICAM-1. LXA<sub>4</sub>, PD1 and RvD1 also enhanced uptake of apoptotic thymocytes, which could contribute to the anti-atherogenic role of this pathway in mice. This process can also be manipulated by diet to govern severity of atherogenic lesions<sup>75,76</sup>. RvE1 is protective in periodontal disease but, unlike LXA<sub>4</sub>, exogenous RvE1 also stimulates bone regeneration in rabbit models of periodontitis<sup>60,77</sup>. In murine models of arthritis, RvD1 and 17-HDHA reduce pain and tissue damage, proving more potent than either steroid or pain treatments<sup>78</sup>.

Unresolved inflammation, epithelial and microvascular injury can lead to excessive fibrosis (Fig. 1) that impairs organ function. Leukotrienes are profibrotic, and in people with scleroderma interstitial lung disease, LXA<sub>4</sub> is present in the lungs at levels that seem unable to counter-regulate profibrotic factors<sup>79</sup>. In animals, an exogenous aspirin-triggered lipoxin analogue reduces pulmonary fibrosis induced by the antibiotic bleomycin<sup>80</sup>, and both LXA<sub>4</sub> and benzo-LXA<sub>4</sub> reduce renal fibrosis<sup>81</sup>. Exogenous RvE1 and RvD1 protect from renal fibrosis, reducing collagen I and IV,  $\alpha$ -SMA and fibronectin<sup>82</sup>. In addition, exogenous RvD1 reduces pro-inflammatory mediators that are generated in response to cigarette smoke and pulmonary toxicants<sup>83</sup>.

### Organ regeneration and wound healing

In mice, exogenous SPMs have been found to stimulate wound healing in a range of studies. LXA<sub>4</sub> stimulates re-epithelialization of the cornea in a gender-specific fashion in mice<sup>84</sup>; when applied to wounds, RvE1, RvD1 and RvD2 each stimulate murine dermal healing, reducing neutrophilic infiltration and stimulating re-epithelialization<sup>85</sup>; RvD1 and RvD2 also stimulate wound healing in diabetic mice<sup>21,86</sup>. Macrophages have a role in wound healing and organ regeneration, and the macrophage-derived maresin pathway stimulates tissue regeneration. This pathway (Fig. 2) is present in the flatworm *Dugesia tigrina*, in which RvE1 and MaR1 were each found to reduce regeneration times (speed of head-segment regrowth)<sup>23</sup>. Given the importance of tissue regeneration in trauma and infection, regulation of resolution programs could hold promise as a therapy.

### SPMs in adaptive immunity

Lymphoid tissue such as mouse spleen produces RvD1, 17-HDHA, PD1 (ref. 87) and LXA<sub>4</sub> (ref. 88) from endogenous sources, which was identified using mass spectrometry, suggesting these products are strategically positioned to act on lymphocytes (Fig. 1). Both exogenous 17-HDHA and RvD1 increase human B-cell immunoglobulin-M and immunoglobulin- $\gamma$ , a response that was not shared by PD1. B-cell differentiation is enhanced by exogenous 17-HDHA towards the CD27<sup>+</sup>CD38<sup>+</sup> antibody-secreting cell phenotype<sup>87</sup>. PD1 is biosynthesized by human T-helper-2-skewed mononuclear cells by a 16(17)-epoxy-protectin intermediate (Fig. 2) and reduces T-cell migration, TNF- $\alpha$  and interferon- $\gamma$ , promoting T-cell apoptosis<sup>89</sup>. Exogenous LXA<sub>4</sub>, RvE1 and PD1 each upregulate expression of the chemokine receptor CCR5 on leukocytes, facilitating their clearance and resolution<sup>90</sup>. Exogenous PD1 reduces CD4<sup>+</sup> T-cell infiltration into cornea<sup>66</sup>, as does RvE1 (ref. 67), in herpes simplex viral infections. Exogenous RvD1 reduces CD11b<sup>+</sup> leukocytes and CD4<sup>+</sup> and CD8<sup>+</sup> T lymphocytes in uveitis<sup>91</sup>. Exogenous RvE1 and RvD1 each regulate T-cell activation in choroid and retina, and are biosynthesized in this tissue, as identified by LC-MS-MS<sup>92</sup>. Exogenous RvE1 induces apoptosis of activated T cells by the induction of 2,3-dioxygenase in dendritic cells, representing a new functional subtype of dendritic cells that have a role in resolution<sup>93</sup>. Exogenous RvE1 has also been shown to reduce numbers of mouse CD4<sup>+</sup> T cells and CD8<sup>+</sup> T cells in atopic dermatitis<sup>94</sup>.

### Neuroinflammation and pain

Mouse and human brains have the capacity to produce resolvins and protectins, as do human microglial cells, in which they reduce cytokine expression<sup>8,95,96</sup>. And the production of SPMs by the brain cells of trout

indicates that they are conserved from fish to humans<sup>88</sup>. In ischaemic mouse brains, immunoreactive resolvins, protectins and their aspirin-triggered forms are produced<sup>97</sup>; in these mice synthetic SPMs are protective, downregulating excess leukocyte infiltration and reducing local neuronal injury, COX-2 induction, and levels of IL-1 $\beta$  and NF- $\kappa$ B. Thus, in the brain, DHA could be a precursor to neuroprotective signalling pathways evoked by ischaemia reflow tissue injury. The DHA product 10,17-dihydroxy-protectin is called neuroprotectin D1 (NPD1) when biosynthesized and acting in neural tissues and retinal epithelial cells owing to its potent actions to reduce neuroinflammation and protect neural cells<sup>98</sup>.

DHA is enriched in the brain, synapses and the retina, in which it is known to have a protective role, but its role as a precursor to mediators in resolution and neuroprotection is still emerging. Synthetic NPD1 has a potent protective role in the nervous system, reducing stress pathways that lead to cell death and increase cell survival, and in several ocular models of important diseases (such as herpes<sup>66</sup> and neovascularization in the eye<sup>99</sup>) NPD1 targets microglia<sup>8,98,99</sup>. Human neutrophils biosynthesize 17R-NPD1/DP1 that is enhanced by aspirin *in vitro*, limiting neutrophil transmigration and enhancing macrophage efferocytosis<sup>37</sup>. Synthetic aspirin-triggered NPD1 reduces brain oedema in penumbra and subcortical lesion size, and improves neurological scores<sup>100</sup>.

In the brains of people with Alzheimer's disease, NPD1 is reduced<sup>96</sup> and the resolution pathway (SPM receptors and products) is diminished<sup>101</sup>. LXA<sub>4</sub> and RvD1 are reduced in cerebrospinal fluid and hippocampus, which correlated with lower scores on the mini-mental state examinations in these patients. These findings provide further evidence that failed resolution of the inflammatory response could contribute to human disease<sup>101</sup>. RvD1 added to macrophages from patients with Alzheimer's *in vitro* reduces the macrophages' pro-inflammatory phenotype and enhances phagocytosis of amyloid- $\beta$ <sup>102</sup>. This is consistent with the suggestion that resolvins promote clearance of amyloid- $\beta$  deposition to reduce inflammation in Alzheimer's. Hence, SPMs might have homeostatic roles in brain and peripheral tissues, whereby each SPM has selective functions to reduce neuroinflammation.

Inflammation can evoke persistent pain. ALX/FPR2 is expressed on spinal astrocytes, and local spinal delivery of synthetic LXA<sub>4</sub>, LXB<sub>4</sub> or their metabolically stable analogues reduces inflammation-induced pain<sup>103</sup>. Many SPMs dampen pain; in mice, they have specific targets of action<sup>104</sup>, as demonstrated with RvE1 and RvD1 for inflammatory pain involving both central and peripheral sites<sup>105</sup>. RvE1 administered intrathecally in mice is more potently analgesic than morphine or a COX-2 inhibitor. The RvE1 receptor ChemR23 is present in dorsal root ganglia, in which synthetic RvE1 regulates phosphorylation of ERK-dependent transient receptor potential vanilloid subtype-1 (TRPV1) inhibition and TNF- $\alpha$ -mediated hyperalgesia, centrally. In postsynaptic neurons, synthetic RvE1 inhibits glutamate and TNF- $\alpha$  stimulation of *N*-methyl-D-aspartic acid (NMDA) receptor and mechanical allodynia<sup>105</sup>. Synthetic RvD1, RvD2 and PD1/NPD1 each reduce pain by the inhibition of specific TRPV channels<sup>106,107</sup>.

MaR1 inhibits TRPV1 in neurons and blocks capsaicin-induced inward current (the half-maximum inhibitory concentration is 0.49 nM), diminishing inflammatory and chemotherapy-evoked neuropathic pain in mice<sup>24</sup>. Both aspirin-triggered RvD1 and 17R-HDHA reduce adjuvant-induced arthritis in rats and associated pain<sup>78</sup>, reduce NF- $\kappa$ B and COX-2 expression in the spinal cord, and, in arthritic joints, reduce TNF- $\alpha$  and IL-1 $\beta$  levels. In addition to leukocytes and microglia, SPM receptors are present on neuronal bodies, nerve terminals (skin and muscle) and synaptic terminals, on which they regulate specific transient receptor potential channels. For example, RvE1–ChemR23 interaction in dorsal root ganglion regulates TRPV1, but not by direct activation of channels like endocannabinoids<sup>104</sup> or other lipids that directly bind transient receptor potential channels; instead, each SPM activates a specific GPCR in the pico-nanomolar range to regulate channels involved in pain signalling.

Direct comparisons between synthetic LXA<sub>4</sub> and aspirin-triggered

RvD1 mechanical hypersensitivity in a rat model of inflammation-induced pain indicate that both effectively reduce hypersensitivity and pro-inflammatory mediators from astrocytes<sup>108</sup>. Local increases in pro-inflammatory mediators can result in cognitive decline, which can occur after major surgery or critical illness and is a major public health concern. Systemic aspirin-triggered RvD1 prophylaxis helps to prevent memory decline in mice that have undergone surgery, protecting them from post-operative neuronal dysfunction<sup>109</sup>. Whether SPM actions in mouse pain models translate to reducing human pain and improving cognition remains of interest.

## Towards human translation

Because resolvins and protectins were identified in mouse exudates, it was essential to establish their biosynthesis by human leukocytes and in human tissues<sup>7,8,95</sup> (Box 2). Using mass spectrometry, RvE1 and RvE2 can be identified in human blood<sup>47,61,110</sup>, provided EPA or DHA is available. RvD1 and RvD2 are found in human plasma and serum<sup>111</sup>. The capacity of murine placenta to form resolvins (RvD1 and RvD2) and protectins (PD1 and 10S,17S-diHDHA<sup>112</sup>, also known as PDx) was confirmed by mass spectrometry, and the levels of each were increased by dietary omega-3 (ref. 113). Strategically, SPMs are also present in human breast milk<sup>114</sup>. LXA<sub>4</sub>, RvD1 and RvE1 were identified in milk of mothers in their first month of lactation using mass spectrometry<sup>114</sup>. These identifications, made possible with LC–MS–MS and the availability of SPMs, could provide opportunities for the rigorous assessment of SPM functional roles in human physiology and their potential for therapy.

Evidence for lipid-mediator class switching in humans has recently been reported<sup>115</sup>. LC–MS–MS lipidomic analysis of venous blood collected after strenuous resistance exercise identified lipid mediators in peripheral blood. Prostanoids in the initial post-exercise recovery phase were temporally followed by leukotrienes and p450-derived eicosanoids (EETs), as well as lipoxins, resolvins and protectins. The NSAID ibuprofen, which is widely used for muscle aches and pains, blocked exercise-induced prostanoids, reduced LTB<sub>4</sub>, and both delayed and diminished the SPM response, as identified by mass spectrometry. This study of resistance exercise in humans illustrates the acute response of pro-inflammatory mediators, presumably from muscle, and the mediators' potential link to resolution programs<sup>115</sup>. However, the amount of resolvins and protectins reported are relative, and additional studies are needed to establish their levels in healthy exercising individuals.

In a 60-patient double-blind trial<sup>116</sup> of infants with eczema, topical 15(R/S)-methyl-LXA<sub>4</sub> relieved severity and improved quality of life without apparent adverse events. In these infants, the lipoxin analogue was as effective as a topical steroid.

## Looking forward

For autacoids, it is when and where that counts, and it is important to emphasize that the first response in acute inflammation is ubiquitous and mounts throughout the body. SPMs are agonists of resolution. Each stimulates cessation of neutrophil influx, efferocytosis and enhances phagocytosis for microbial containment — signs of resolution (Fig. 1). These steps are the defining SPM functions. Each SPM pathway has additional, non-redundant functions on target cell types. At the cellular and molecular levels, SPMs counter-regulate pro-inflammatory mediators (eicosanoids, chemokines, cytokines<sup>33</sup> and adipokines<sup>22</sup>); regulate specific miRNAs<sup>53</sup> and cell traffic; and enhance microbial killing by receptor-mediated mechanisms in animal models *in vivo* and with human neutrophils and macrophages<sup>10,13,21</sup>.

Results from preclinical disease models<sup>10</sup> suggest that treatment of inflammation-associated disease might be possible with SPM agonists that stimulate resolution and protect organs from collateral damage (Box 2). RvE1, MaR1 and NPD1/DP1 are each in clinical development programs. An RvE1 mimetic is in clinical trials for ocular indications (<http://www.auventx.com/auven/products/rx10045.php>), and NPD1/DP1 is in clinical development for neurodegenerative diseases (<http://www.anidapharma.com/lead-molecule.html>); given their ability to



regulate inflammation resolution without immunosuppression, it is hopeful they will become therapies.

Because identifying SPMs in human tissues relies on LC–MS–MS-based approaches and internal standards that have only recently become available, at this point relatively few studies have demonstrated the presence of SPMs in human tissues (such as blood, milk, and adipose and brain tissue). Hence, evidence for SPM formation in humans is at a very early stage, and their functional importance in human health and diseases remains to be established. Whether SPMs have physiological actions in target tissues in humans, now that we know that they are produced in levels that show potent selective actions in animals<sup>10,19</sup>, can now be addressed (Box 2) — along with the impact of anti-inflammatories on SPM pathways — using LC–MS–MS-based lipid-mediator–SPM profiling of human tissues. The SPMs produced by human apoptotic neutrophils and macrophages can be identified using LC–MS–MS profiling<sup>13</sup> without omega-3 EFA supplementation; it is therefore now possible, with the sensitivity of LC–MS–MS, to determine individual human SPM profiles (personalized lipid mediator–SPM metabolomics). This personalized metabolomics can be used to determine the relationship to resolution mechanisms. It is also important to determine, in healthy individuals, the omega-3 EFA supplementation and doses that might increase or diminish SPMs within specific tissues. Importantly, it is crucial to assess whether human diseases that are characterized by excessive inflammation result from failed resolution mechanisms because of the absence of specific SPM pathways and whether these mechanisms can be rescued either by EPA or DHA, or by therapeutic versions of SPM mimetics. Given that resolution of inflammation is a fundamental process in all organs, new approaches for stimulating resolution and using SPMs as pathway markers could soon be possible, as could determining the relationship of SPMs to nutrition in humans.

SPMs have emerged from animal models as potential regulators in physiological pathways of resolution and unresolved inflammation that can affect infection, pain, obesity, organ protection<sup>21</sup> and inflammatory diseases (Box 2) beyond the roles of their omega-3 EFA precursors in intermediary metabolism and membrane dynamics. Identification of SPM bioactive metabolomes and an appreciation that exudates can drive resolution, partly through SPMs, sets a new terrain to evaluate resolution physiology and pharmacology, in which SPMs are crucial as chemical signals for catabasis and host defence.

**Note added in proof:** A paper recently appeared online while the current Review was in press reporting the quantification of pro-resolving mediators using a new approach in human tissues and their functional relationship to known eicosanoids<sup>121</sup>. ■

Received 18 December 2013; accepted 24 March 2014.

- Flower, R. J. Prostaglandins, bioassay and inflammation. *Br. J. Pharmacol.* **147**, S182–S192 (2006).
  - Samuelsson, B. Role of basic science in the development of new medicines: examples from the eicosanoid field. *J. Biol. Chem.* **287**, 10070–10080 (2012).
  - Dinarello, C. A., Simon, A. & van der Meer, J. W. Treating inflammation by blocking interleukin-1 in a broad spectrum of diseases. *Nature Rev. Drug Discov.* **11**, 633–652 (2012).
  - Serhan, C. N. & Savill, J. Resolution of inflammation: the beginning programs the end. *Nature Immunol.* **6**, 1191–1197 (2005).
  - Maderna, P. & Godson, C. Lipoxins: resolutionary road. *Br. J. Pharmacol.* **158**, 947–959 (2009).
  - Tabas, I. & Glass, C. K. Anti-inflammatory therapy in chronic disease: challenges and opportunities. *Science* **339**, 166–172 (2013).
  - Serhan, C. N. *et al.* Novel functional sets of lipid-derived mediators with antiinflammatory actions generated from omega-3 fatty acids via cyclooxygenase 2-nonsteroidal antiinflammatory drugs and transcellular processing. *J. Exp. Med.* **192**, 1197–1204 (2000).
  - Serhan, C. N. *et al.* Resolvins: a family of bioactive products of omega-3 fatty acid transformation circuits initiated by aspirin treatment that counter pro-inflammation signals. *J. Exp. Med.* **196**, 1025–1037 (2002).
  - Lands, W. E. M. *Fish, Omega-3 and Human Health* 2nd edn (AOCS Press, 2005).
  - Serhan, C. N. & Chiang, N. Resolution phase lipid mediators of inflammation: agonists of resolution. *Curr. Opin. Pharmacol.* **13**, 632–640 (2013).
  - Malawista, S. E., de Boisleury Chevance, A., van Damme, J. & Serhan, C. N. Tonic inhibition of chemotaxis in human plasma. *Proc. Natl Acad. Sci. USA* **105**, 17949–17954 (2008).
  - Fullerton, J. N., O'Brien, A. J. & Gilroy, D. W. Lipid mediators in immune dysfunction after severe inflammation. *Trends Immunol.* **35**, 12–21 (2014).
  - Dalli, J. & Serhan, C. N. Specific lipid mediator signatures of human phagocytes: microparticles stimulate macrophage efferocytosis and pro-resolving mediators. *Blood* **120**, e60–e72 (2012).
  - Serhan, C. N. The resolution of inflammation: the devil in the flask and in the details. *FASEB J.* **25**, 1441–1448 (2011).
  - Bailey, R. L., Gahche, J. J., Miller, P. E., Thomas, P. R. & Dwyer, J. T. Why US adults use dietary supplements. *JAMA Intern. Med.* **173**, 355–361 (2013).
  - Yates, C. M., Calder, P. C. & Rainger, G. E. Pharmacology and therapeutics of omega-3 polyunsaturated fatty acids in chronic inflammatory disease. *Pharmacol. Ther.* **141**, 272–282 (2014).
- A thoughtful review of the use of omega-3 essential fatty acids in inflammatory diseases and their potential in supplementation, dietary manipulation and pharmacology.**
- Serhan, C. N. & Petasis, N. A. Resolvins and protectins in inflammation-resolution. *Chem. Rev.* **111**, 5922–5943 (2011).
  - Schwab, J. M., Chiang, N., Arita, M. & Serhan, C. N. Resolvin E1 and protectin D1 activate inflammation-resolution programmes. *Nature* **447**, 869–874 (2007).
  - Miyahara, T. *et al.* D-series resolvins attenuate vascular smooth muscle cell activation and neointimal hyperplasia following vascular injury. *FASEB J.* **27**, 2220–2232 (2013).
  - Levy, B. D. & Serhan, C. N. Resolution of acute inflammation in the lung. *Annu. Rev. Physiol.* **76**, 467–492 (2014).
  - Spite, M., Claria, J. & Serhan, C. N. Resolvins, specialized proresolving lipid mediators, and their potential roles in metabolic diseases. *Cell Metab.* **19**, 21–36 (2014).
  - Dalli, J. *et al.* Resolvin D3 and aspirin-triggered resolvin D3 are potent immunoresolvents. *Chem. Biol.* **20**, 188–201 (2013).
  - Serhan, C. N. *et al.* Macrophage pro-resolving mediator maresin 1 stimulates tissue regeneration and controls pain. *FASEB J.* **26**, 1755–1765 (2012).
  - Dalli, J. *et al.* The novel 13S,14S-epoxy-maresin is converted by human macrophages to maresin1 (MaR1), inhibits leukotriene A4 hydrolase (LTA4H), and shifts macrophage phenotype. *FASEB J.* **27**, 2573–2583 (2013).
  - Biswas, S. K. & Mantovani, A. Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. *Nature Immunol.* **11**, 889–896 (2010).
  - Levy, B. D., Clish, C. B., Schmidt, B., Gronert, K. & Serhan, C. N. Lipid mediator class switching during acute inflammation: signals in resolution. *Nature Immunol.* **2**, 612–619 (2001).
  - Morris, T. *et al.* Effects of low-dose aspirin on acute inflammatory responses in humans. *J. Immunol.* **183**, 2089–2096 (2009).
  - Takano, T., Clish, C. B., Gronert, K., Petasis, N. & Serhan, C. N. Neutrophil-mediated changes in vascular permeability are inhibited by topical application of aspirin-triggered 15-epi-lipoxin A4 and novel lipoxin B4 stable analogues. *J. Clin. Invest.* **101**, 819–826 (1998).
  - Chan, M. M.-Y. & Moore, A. R. Resolution of inflammation in murine autoimmune arthritis is disrupted by cyclooxygenase-2 inhibition and restored by prostaglandin E2-mediated lipoxin A4 production. *J. Immunol.* **184**, 6418–6426 (2010).
  - Bannenberg, G. L. *et al.* Molecular circuits of resolution: Formation and actions of resolvins and protectins. *J. Immunol.* **174**, 4345–4355 (2005).
  - Lucas, C. D. *et al.* Downregulation of Mcl-1 has anti-inflammatory pro-resolution effects and enhances bacterial clearance from the lung. *Mucosal Immunol.* <http://dx.doi.org/10.1038/mi.2013.102> (2013).
  - Navarro-Xavier, R. A. *et al.* A new strategy for the identification of novel molecules with targeted proresolution of inflammation properties. *J. Immunol.* **184**, 1516–1525 (2010).
  - Chiang, N. *et al.* Infection regulates pro-resolving mediators that lower antibiotic requirements. *Nature* **484**, 524–528 (2012).
  - Spite, M. *et al.* Resolvin D2 is a potent regulator of leukocytes and controls microbial sepsis. *Nature* **461**, 1287–1291 (2009).
  - Birnbaum, Y. *et al.* Augmentation of myocardial production of 15-epi-lipoxin-A4 by pioglitazone and atorvastatin in the rat. *Circulation* **114**, 929–935 (2006).
- This is the first description of 15-epi-LXA<sub>4</sub> biosynthesis by statins *in vivo* in rats using a novel pathway and phosphorylated enzymes in the biosynthesis of 15-epi-LXA<sub>4</sub> that mimic the biosynthetic capacity of acetylated COX-2.**
- Cooray, S. N. *et al.* Ligand-specific conformational change of the G-protein-coupled receptor ALX/FPR2 determines proresolving functional responses. *Proc. Natl Acad. Sci. USA* **110**, 18232–18237 (2013).
- This article reports exciting results indicating that heterodimers and homodimers of ALX/FPR2 signal different responses by activation of distinct intracellular signalling mechanisms.**
- Serhan, C. N. *et al.* Novel proresolving aspirin-triggered DHA pathway. *Chem. Biol.* **18**, 976–987 (2011).
  - Morris, T. *et al.* Dichotomy in duration and severity of acute inflammatory responses in humans arising from differentially expressed proresolution pathways. *Proc. Natl Acad. Sci. USA* **107**, 8842–8847 (2010).
  - Brancaleone, V. *et al.* A vasculo-protective circuit centered on lipoxin A4 and aspirin-triggered 15-epi-lipoxin A4 operative in murine microcirculation. *Blood* **122**, 608–617 (2013).
  - Norling, L. V. *et al.* Cutting edge: humanized nano-proresolving medicines mimic inflammation-resolution and enhance wound healing. *J. Immunol.* **186**, 5543–5547 (2011).
  - Stables, M. J. *et al.* Transcriptomic analyses of murine resolution-phase macrophages. *Blood* **118**, e192–e208 (2011).



42. Miki, Y. *et al.* Lymphoid tissue phospholipase A2 group IID resolves contact hypersensitivity by driving antiinflammatory lipid mediators. *J. Exp. Med.* **210**, 1217–1234 (2013).  
**This article reports the identification of a secretory phospholipase A2 activated during contact hypersensitive reactions that is specifically involved in the generation of resolvin D1 and protectin D1.**
43. Levy, B. D. *et al.* Protectin D1 is generated in asthma and dampens airway inflammation and hyper-responsiveness. *J. Immunol.* **178**, 496–502 (2007).
44. Miyata, J. *et al.* Dysregulated synthesis of protectin D1 in eosinophils from patients with severe asthma. *J. Allergy Clin. Immunol.* **131**, 353–360 (2013).
45. Yamada, T. *et al.* Eosinophils promote resolution of acute peritonitis by producing proresolving mediators in mice. *FASEB J.* **25**, 561–568 (2011).
46. Isobe, Y. *et al.* Stereochemical assignment and anti-inflammatory properties of the omega-3 lipid mediator resolvin E3. *J. Biochem.* **153**, 355–360 (2013).
47. Arita, M. *et al.* Stereochemical assignment, anti-inflammatory properties, and receptor for the omega-3 lipid mediator resolvin E1. *J. Exp. Med.* **201**, 713–722 (2005).
48. Ohira, T. *et al.* Resolvin E1 receptor activation signals phosphorylation and phagocytosis. *J. Biol. Chem.* **285**, 3451–3461 (2010).
49. El Kebir, D., Gjørstrup, P. & Filep, J. G. Resolvin E1 promotes phagocytosis-induced neutrophil apoptosis and accelerates resolution of pulmonary inflammation. *Proc. Natl Acad. Sci. USA* **109**, 14983–14988 (2012).
50. Krishnamoorthy, S., Recchiuti, A., Chiang, N., Fredman, G. & Serhan, C. N. Resolvin D1 receptor stereoselectivity and regulation of inflammation and pro-resolving microRNAs. *Am. J. Pathol.* **180**, 2018–2027 (2012).
51. Norling, L. V., Dall, J., Flower, R. J., Serhan, C. N. & Perretti, M. Resolvin D1 limits polymorphonuclear leukocyte recruitment to inflammatory loci: receptor dependent actions. *Arterioscler. Thromb. Vasc. Biol.* **32**, 1970–1978 (2012).
52. Fredman, G., Li, Y., Dall, J., Chiang, N. & Serhan, C. N. Self-limited versus delayed resolution of acute inflammation: temporal regulation of pro-resolving mediators and microRNA. *Sci. Rep.* **2**, 639 (2012).
53. Li, Y. *et al.* Plasticity of leukocytic exudates in resolving acute inflammation is regulated by microRNA and proresolving mediators. *Immunity* **39**, 885–898 (2013).
54. Li, D. *et al.* Resolvin D1 and aspirin-triggered resolvin D1 regulate histamine-stimulated conjunctival goblet cell secretion. *Mucosal Immunol.* <http://dx.doi.org/10.1038/mi.2013.7> (2013).
55. Nelson, J. *et al.* The ALX/FPR2 receptor for RvD1 is expressed and functional in salivary glands. *Am. J. Physiol. Cell Physiol.* **306**, C178–C185 (2014).
56. Jones, C. N. *et al.* Microfluidic chambers for monitoring leukocyte trafficking and humanized nano-proresolving medicines interactions. *Proc. Natl Acad. Sci. USA* **109**, 20560–20565 (2012).
57. Kasuga, K. *et al.* Rapid appearance of resolvin precursors in inflammatory exudates: novel mechanisms in resolution. *J. Immunol.* **181**, 8677–8687 (2008).
58. Simiele, F. *et al.* Transcriptional regulation of the human FPR2/ALX gene: evidence of a heritable genetic variant that impairs promoter activity. *FASEB J.* **26**, 1323–1333 (2012).
59. Serhan, C. N. *et al.* Reduced inflammation and tissue damage in transgenic rabbits overexpressing 15-lipoxygenase and endogenous anti-inflammatory lipid mediators. *J. Immunol.* **171**, 6856–6865 (2003).
60. Hasturk, H. *et al.* Resolvin E1 regulates inflammation at the cellular and tissue level and restores tissue homeostasis *in vivo*. *J. Immunol.* **179**, 7021–7029 (2007).
61. Oh, S. F., Pillai, P. S., Recchiuti, A., Yang, R. & Serhan, C. N. Pro-resolving actions and stereoselective biosynthesis of 18S E-series resolvins in human leukocytes and murine inflammation. *J. Clin. Invest.* **121**, 569–581 (2011).
62. Walker, J. *et al.* Lipoxin A4 increases survival by decreasing systemic inflammation and bacterial load in sepsis. *Shock* **36**, 410–416 (2011).
63. Prescott, D. & McKay, D. M. Aspirin-triggered lipoxin enhances macrophage phagocytosis of bacteria while inhibiting inflammatory cytokine production. *Am. J. Physiol. Gastrointest. Liver Physiol.* **301**, G487–G497 (2011).
64. Divangahi, M. *et al.* *Mycobacterium tuberculosis* evades macrophage defenses by inhibiting plasma membrane repair. *Nature Immunol.* **10**, 899–906 (2009).
65. Tobin, D. M. *et al.* Host genotype-specific therapies can optimize the inflammatory response to mycobacterial infections. *Cell* **148**, 434–446 (2012).  
**This systematic analysis of tuberculosis in infections in zebrafish identified the host gradation response to producing excess LTB<sub>4</sub> or excess LXA<sub>4</sub>, each having different outcomes for the host and the clearance of tuberculosis infections.**
66. Rajasagi, N. K., Reddy, P. B., Mulik, S., Gjørstrup, P. & Rouse, B. T. Neuroprotectin D1 reduces the severity of herpes simplex virus-induced corneal immunopathology. *Invest. Ophthalmol. Vis. Sci.* **54**, 6269–6279 (2013).
67. Rajasagi, N. K. *et al.* Controlling herpes simplex virus-induced ocular inflammatory lesions with the lipid-derived mediator resolvin E1. *J. Immunol.* **186**, 1735–1746 (2011).  
**This is an exciting first report demonstrating that resolvin E1 protects the eye from Herpes simplex virus infection and stimulates clearance of the virus.**
68. Cilloniz, C. *et al.* Lethal dissemination of H5N1 influenza virus is associated with dysregulation of inflammation and lipoxin signaling in a mouse model of infection. *J. Virol.* **84**, 7613–7624 (2010).
69. Morita, M. *et al.* The lipid mediator protectin D1 inhibits influenza virus replication and improves severe influenza. *Cell* **153**, 112–125 (2013).
70. Tam, V. C. *et al.* Lipidomic profiling of influenza infection identifies mediators that induce and resolve inflammation. *Cell* **154**, 213–227 (2013).
71. Haas-Stapleton, E. J. *et al.* *Candida albicans* modulates host defense by biosynthesizing the pro-resolving mediator resolvin E1. *PLoS ONE* **2**, e1316 (2007).
72. Baillie, J. K. & Digard, P. Influenza—time to target the host? *N. Engl. J. Med.* **369**, 191–193 (2013).
73. Pouliot, M., Clish, C. B., Petasis, N. A., Van Dyke, T. E. & Serhan, C. N. Lipoxin A4 analogues inhibit leukocyte recruitment to *Porphyromonas gingivalis*: a role for cyclooxygenase-2 and lipoxins in periodontal disease. *Biochemistry* **39**, 4761–4768 (2000).
74. Shen, J. *et al.* Macrophage-mediated 15-lipoxygenase expression protects against atherosclerosis development. *J. Clin. Invest.* **98**, 2201–2208 (1996).
75. Merched, A. J., Ko, K., Gotlinger, K. H., Serhan, C. N. & Chan, L. Atherosclerosis: evidence for impairment of resolution of vascular inflammation governed by specific lipid mediators. *FASEB J.* **22**, 3595–3606 (2008).
76. Merched, A. J., Serhan, C. N. & Chan, L. Nutritional disruption of inflammation-resolution homeostasis and atherogenesis. *J. Nutrigenet. Nutrigenomics* **4**, 12–24 (2011).
77. Hasturk, H. *et al.* RvE1 protects from local inflammation and osteoclast mediated bone destruction in periodontitis. *FASEB J.* **20**, 401–403 (2006).
78. Lima-Garcia, J. F. *et al.* The precursor of resolvin D series and aspirin-triggered resolvin D1 display anti-hyperalgesic properties in adjuvant-induced arthritis in rats. *Br. J. Pharmacol.* **164**, 278–293 (2011).  
**The authors of this article demonstrate that 17-HDHA and aspirin-triggered RvD1 have potent protective actions in adjuvant-induced arthritis and reduce pain in this model of arthritis.**
79. Kowal-Bielecka, O., Kowal, K., Distler, O. & Gay, S. Mechanisms of disease: leukotrienes and lipoxins in scleroderma lung disease—insights and potential therapeutic implications. *Nature Clin. Pract. Rheumatol.* **3**, 43–51 (2007).
80. Martins, V. *et al.* ATLa, an aspirin-triggered lipoxin A4 synthetic analog, prevents the inflammatory and fibrotic effects of bleomycin-induced pulmonary fibrosis. *J. Immunol.* **182**, 5374–5381 (2009).
81. Börgeson, E. *et al.* Lipoxin A(4) and benzo-lipoxin A(4) attenuate experimental renal fibrosis. *FASEB J.* **25**, 2967–2979 (2011).  
**This article reported that expediting resolution and counter-regulation of pro-inflammatory mediators by administration of LXA<sub>4</sub> or its analogues can prevent organ fibrosis and renal fibrosis.**
82. Qu, X. *et al.* Resolvins E1 and D1 inhibit interstitial fibrosis in the obstructed kidney via inhibition of local fibroblast proliferation. *J. Pathol.* **228**, 506–519 (2012).
83. Hsiao, H. M. *et al.* A novel anti-inflammatory and pro-resolving role for resolvin D1 in acute cigarette smoke-induced lung inflammation. *PLoS ONE* **8**, e58258 (2013).
84. Wang, S. B. *et al.* Estrogen negatively regulates epithelial wound healing and protective lipid mediator circuits in the cornea. *FASEB J.* **26**, 1506–1516 (2012).
85. Menon, R. *The Effect of Resolvins on Dermal Wound Healing*. MS thesis, Rutgers Univ. (2012).
86. Tang, Y. *et al.* Proresolution therapy for the treatment of delayed healing of diabetic wounds. *Diabetes* **62**, 618–627 (2013).  
**In this article, the authors demonstrate that pro-resolving mediators such as RvD1 can expedite healing of wounds encountered in diabetes.**
87. Ramon, S., Gao, F., Serhan, C. N. & Phipps, R. P. Specialized proresolving mediators enhance human B cell differentiation to antibody-secreting cells. *J. Immunol.* **189**, 1036–1042 (2012).
88. Hong, S. *et al.* Resolvin D1, protectin D1, and related docosahexaenoic acid-derived products: analysis via electrospray/low energy tandem mass spectrometry based on spectra and fragmentation mechanisms. *J. Am. Soc. Mass Spectrom.* **18**, 128–144 (2007).
89. Ariel, A. *et al.* The docosatriene protectin D1 is produced by TH<sub>2</sub> skewing and promotes human T cell apoptosis via lipid raft clustering. *J. Biol. Chem.* **280**, 43079–43086 (2005).
90. Ariel, A. *et al.* Apoptotic neutrophils and T cells sequester chemokines during immune response resolution via modulation of CCR5 expression. *Nature Immunol.* **7**, 1209–1216 (2006).
91. Settimio, R., Clara, D. F., Franca, F., Francesca, S. & Michele, D. Resolvin D1 reduces the immunoinflammatory response of the rat eye following uveitis. *Mediators Inflamm.* **2012**, 318621 (2012).
92. Tian, H., Lu, Y., Sherwood, A. M., Hongqian, D. & Hong, S. Resolvins E1 and D1 in choroid-retinal endothelial cells and leukocytes: biosynthesis and mechanisms of anti-inflammatory actions. *Invest. Ophthalmol. Vis. Sci.* **50**, 3613–3620 (2009).  
**The authors of this report demonstrate the endogenous biosynthesis of RvE1 and RvD1 in the retina and the role of cell-cell interactions in this organ.**
93. Vassiliou, E. K., Kesler, O. M., Tadros, J. H. & Ganea, D. Bone marrow-derived dendritic cells generated in the presence of resolvin E1 induce apoptosis of activated CD4<sup>+</sup> T cells. *J. Immunol.* **181**, 4534–4544 (2008).
94. Kim, T. H., Kim, G. D., Jin, Y. H., Park, Y. S. & Park, C. S. Omega-3 fatty acid-derived mediator, Resolvin E1, ameliorates 2,4-dinitrofluorobenzene-induced atopic dermatitis in NC/Nga mice. *Int. Immunopharmacol.* **14**, 384–391 (2012).
95. Hong, S., Gronert, K., Devchand, P., Moussignac, R.-L. & Serhan, C. N. Novel docosatrienes and 17S-resolvins generated from docosahexaenoic acid in murine brain, human blood and glial cells: autacoids in anti-inflammation. *J. Biol. Chem.* **278**, 14677–14687 (2003).
96. Lukiw, W. J. *et al.* A role for docosahexaenoic acid-derived neuroprotectin D1 in neural cell survival and Alzheimer disease. *J. Clin. Invest.* **115**, 2774–2783 (2005).

97. Marcheselli, V. L. *et al.* Novel docosanoids inhibit brain ischemia-reperfusion-mediated leukocyte infiltration and pro-inflammatory gene expression. *J. Biol. Chem.* **278**, 43807–43817 (2003).  
**This is an important first report of the production and actions of NPD1 in the mouse brain and its neural function by reducing inflammation.**
98. Bazan, N. G., Calandria, J. M. & Serhan, C. N. Rescue and repair during photoreceptor cell renewal mediated by docosahexaenoic acid-derived neuroprotectin D1. *J. Lipid Res.* **51**, 2018–2031 (2010).  
**An authoritative, critical review of the bioactivity and actions of NPD1 in the eye and brain.**
99. Sheets, K. G. *et al.* Microglial ramification and redistribution concomitant with the attenuation of choroidal neovascularization by neuroprotectin D1. *Mol. Vis.* **19**, 1747–1759 (2013).  
**The article is the first description of NPD1 action in regulating neovascularization by targeting microglia.**
100. Bazan, N. G. *et al.* Novel aspirin-triggered neuroprotectin D1 attenuates cerebral ischemic injury after experimental stroke. *Exp. Neurol.* **236**, 122–130 (2012).
101. Wang, X. *et al.* Resolution of inflammation is altered in Alzheimer's disease. *Alzheimers Dement.* <http://dx.doi.org/10.1016/j.jalz.2013.12.024> (2014).
102. Mizwicki, M. T. *et al.* 1 $\alpha$ ,25-Dihydroxyvitamin D3 and resolvin D1 retune the balance between amyloid- $\beta$  phagocytosis and inflammation in Alzheimer's disease patients. *J. Alzheimers Dis.* **34**, 155–170 (2013).
103. Svensson, C. I., Zattoni, M. & Serhan, C. N. Lipoxins and aspirin-triggered lipoxin stop inflammatory pain processing. *J. Exp. Med.* **204**, 245–252 (2007).
104. Piomelli, D. & Sasso, O. Peripheral gating of pain signals by endogenous lipid mediators. *Nature Neurosci.* **17**, 164–174 (2014).
105. Xu, Z.-Z. *et al.* Resolvins RvE1 and RvD1 attenuate inflammatory pain via central and peripheral actions. *Nature Med.* **16**, 592–597 (2010).
106. Bang, S., Yoo, S., Yang, T. J., Cho, H. & Hwang, S. W. 17(R)-resolvin D1 specifically inhibits transient receptor potential ion channel vanilloid 3 leading to peripheral antinociception. *Br. J. Pharmacol.* **165**, 683–692 (2012).
107. Park, C. K. *et al.* Resolving TRPV1 and TNF- $\alpha$ -mediated spinal cord synaptic plasticity and inflammatory pain with neuroprotectin D1. *J. Neurosci.* **31**, 15072–15085 (2011).
108. Abdelmoaty, S. *et al.* Spinal actions of lipoxin A4 and 17(R)-resolvin D1 attenuate inflammation-induced mechanical hypersensitivity and spinal TNF release. *PLoS ONE* **8**, e75543 (2013).
109. Terrando, N. *et al.* Aspirin-triggered resolvin D1 prevents surgery-induced cognitive decline. *FASEB J.* **27**, 3564–3571 (2013).  
**This article reports an exciting discovery demonstrating that RvD1 counter-regulates pro-inflammatory mediators produced during surgery-induced cognitive decline.**
110. Psychogios, N. *et al.* The human serum metabolome. *PLoS ONE* **6**, e16957 (2011).
111. Mas, E., Croft, K. D., Zahra, P., Barden, A. & Mori, T. A. Resolvins D1, D2, and other mediators of self-limited resolution of inflammation in human blood following n-3 fatty acid supplementation. *Clin. Chem.* **58**, 1476–1484 (2012).
112. Serhan, C. N. *et al.* Anti-inflammatory actions of neuroprotectin D1/protectin D1 and its natural stereoisomers: assignments of dihydroxy-containing docosatrienes. *J. Immunol.* **176**, 1848–1859 (2006).
113. Jones, M. L. *et al.* Maternal dietary omega-3 fatty acid intake increases resolvin and protectin levels in the rat placenta. *J. Lipid Res.* **54**, 2247–2254 (2013).  
**This article is an important contribution demonstrating that resolvins and protectins are present in placenta and that their levels can be substantially increased with dietary supplementation.**
114. Weiss, G. A. *et al.* High levels of anti-inflammatory and pro-resolving lipid mediators lipoxins and resolvins and declining docosahexaenoic acid levels in human milk during the first month of lactation. *Lipids Health Dis.* **12**, 89 (2013).
115. Markworth, J. F. *et al.* Human inflammatory and resolving lipid mediator responses to resistance exercise and ibuprofen treatment. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **305**, R1281–R1296 (2013).  
**This article reports exciting new results indicating that strenuous exercise activates acute inflammation and pro-inflammatory eicosanoids, which then transition in humans to pro-resolving mediators present in peripheral blood following a time course consistent with lipid-mediator class switching and resolution of exercise-induced muscle stress or inflammation.**
116. Wu, S. H., Chen, X. Q., Liu, B., Wu, H. J. & Dong, L. Efficacy and safety of 15(R/S)-methyl-lipoxin A<sub>4</sub> in topical treatment of infantile eczema. *Br. J. Dermatol.* **168**, 172–178 (2013).  
**This article reports a paediatric clinical trial that was the first to demonstrate that the topical addition of an aspirin-triggered LXA<sub>4</sub> stable analogue is safe and effective in reducing infantile eczema.**
117. Raatz, S. K. *et al.* Baking reduces prostaglandin, resolvin, and hydroxy-fatty acid content of farm-raised Atlantic salmon (*Salmo salar*). *J. Agric. Food Chem.* **59**, 11278–11286 (2011).
118. Clària, J., Nguyen, B. T., Madenci, A., Ozaki, C. K. & Serhan, C. N. Diversity of lipid mediators in human adipose tissue depots. *Am. J. Physiol. Cell Physiol.* **304**, C1141–C1149 (2013).
119. Prüss, H. *et al.* Proresolution lipid mediators in multiple sclerosis - differential, disease severity-dependent synthesis — a clinical pilot trial. *PLoS ONE* **8**, e55859 (2013).
120. Giera, M. *et al.* Lipid and lipid mediator profiling of human synovial fluid in rheumatoid arthritis patients by means of LC-MS/MS. *Biochim. Biophys. Acta* **1821**, 1415–1424 (2012).
121. Colas, R., Shinohara, M., Dalli, J., Chiang, N. & Serhan, C. N. Identification and signature profiles for pro-resolving and inflammatory lipid mediators in human tissue. *Am. J. Physiol. Cell Physiol.* <http://dx.doi.org/10.1152/ajpcell.00024.2014> (2 April, 2014).

**Acknowledgments** The author thanks S. Orr, J. Dalli and N. Chiang for their critical reading of this manuscript and grants from the US National Institutes of Health (R01GM038765 and P01GM095467) and the Mérieux Foundation (France) for support of the author's research. I also thank our collaborators and investigators contributing to this area whose publications were not cited due to size limitations.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The author declares competing financial interests see [go.nature.com/4lixkv](http://go.nature.com/4lixkv) for details. Readers are welcome to comment on the online version of this paper at [go.nature.com/4lixkv](http://go.nature.com/4lixkv). Correspondence should be addressed to C.N.S. ([cnserhan@zeus.bwh.harvard.edu](mailto:cnserhan@zeus.bwh.harvard.edu)).

# Accurate design of co-assembling multi-component protein nanomaterials

Neil P. King<sup>1,2\*</sup>, Jacob B. Bale<sup>1,3\*</sup>, William Sheffler<sup>1\*</sup>, Dan E. McNamara<sup>4</sup>, Shane Gonen<sup>1,5</sup>, Tamir Gonen<sup>5</sup>, Todd O. Yeates<sup>4,6,7</sup> & David Baker<sup>1,2,8</sup>

**The self-assembly of proteins into highly ordered nanoscale architectures is a hallmark of biological systems. The sophisticated functions of these molecular machines have inspired the development of methods to engineer self-assembling protein nanostructures; however, the design of multi-component protein nanomaterials with high accuracy remains an outstanding challenge. Here we report a computational method for designing protein nanomaterials in which multiple copies of two distinct subunits co-assemble into a specific architecture. We use the method to design five 24-subunit cage-like protein nanomaterials in two distinct symmetric architectures and experimentally demonstrate that their structures are in close agreement with the computational design models. The accuracy of the method and the number and variety of two-component materials that it makes accessible suggest a route to the construction of functional protein nanomaterials tailored to specific applications.**

The unique functional opportunities afforded by protein self-assembly range from the dynamic cellular scaffolding provided by cytoskeletal proteins to the encapsulation, protection and delivery of viral genomes to new host cells by virus capsids. Although natural assemblies can be repurposed to perform new functions<sup>1,2</sup>, this strategy is limited to the structures of existing proteins, which may not be suited to a given application. To overcome this limitation, methods for designing novel self-assembling proteins are of considerable interest<sup>3–6</sup>. The central challenge in designing self-assembling proteins is to encode the information necessary to direct assembly in the structures of the protein building blocks. Although the complexity and irregularity of protein structures resulted in slow initial progress in this area, advances in computational protein design algorithms and new approaches such as metal-mediated assembly have recently yielded exciting results<sup>6–16</sup>. Despite these advances, the self-assembling protein structures designed so far have been relatively simple, and continued improvements in design strategies are needed in order to enable the practical design of functional materials.

The level of structural complexity available to self-assembled nanomaterials generally increases with the number of unique molecular components used to construct the material. This is illustrated by DNA nanotechnology, in which specific and directional interactions between hundreds of distinct DNA strands allow the construction of nanoscale objects with essentially arbitrary structures<sup>17–20</sup>. In contrast, designing well-ordered multi-component protein nanomaterials has remained a significant challenge. Multiple distinct intermolecular contacts are necessary to drive the assembly of such materials<sup>3,4,8,11,21</sup>, and programming new, geometrically precise interactions between proteins is difficult. Compared to homo-oligomers, multi-component protein nanomaterials offer several advantages: a wider range of possible structures due to their construction from combinations of building blocks, greater control over the timing of assembly, and enhanced modularity through independently addressable subunits. Although multi-component protein assemblies have recently been generated using disulphide bonds<sup>14,22</sup>, flexible genetic linkers<sup>11,15,22</sup>, or stereotyped coiled-coil interactions to drive

assembly<sup>14,15</sup>, the flexibility of these relatively minimal linkages has generally resulted in materials that are somewhat polydisperse. Most natural protein assemblies, on the other hand, are constructed from protein–protein interfaces involving many contacts distributed over large interaction surfaces that serve to precisely define the positions of the subunits relative to each other<sup>23,24</sup>. Advances in computational protein modelling and design algorithms have recently made it possible to design such interfaces<sup>25–29</sup> and thereby direct the formation of novel self-assembling protein nanomaterials with atomic-level accuracy<sup>7,9,10</sup>, but the methods reported so far have been limited to the design of materials comprising only a single type of molecular building block. Here we expand the structural and functional range of designed protein materials with a general computational method for designing two-component co-assembling protein nanomaterials with high accuracy.

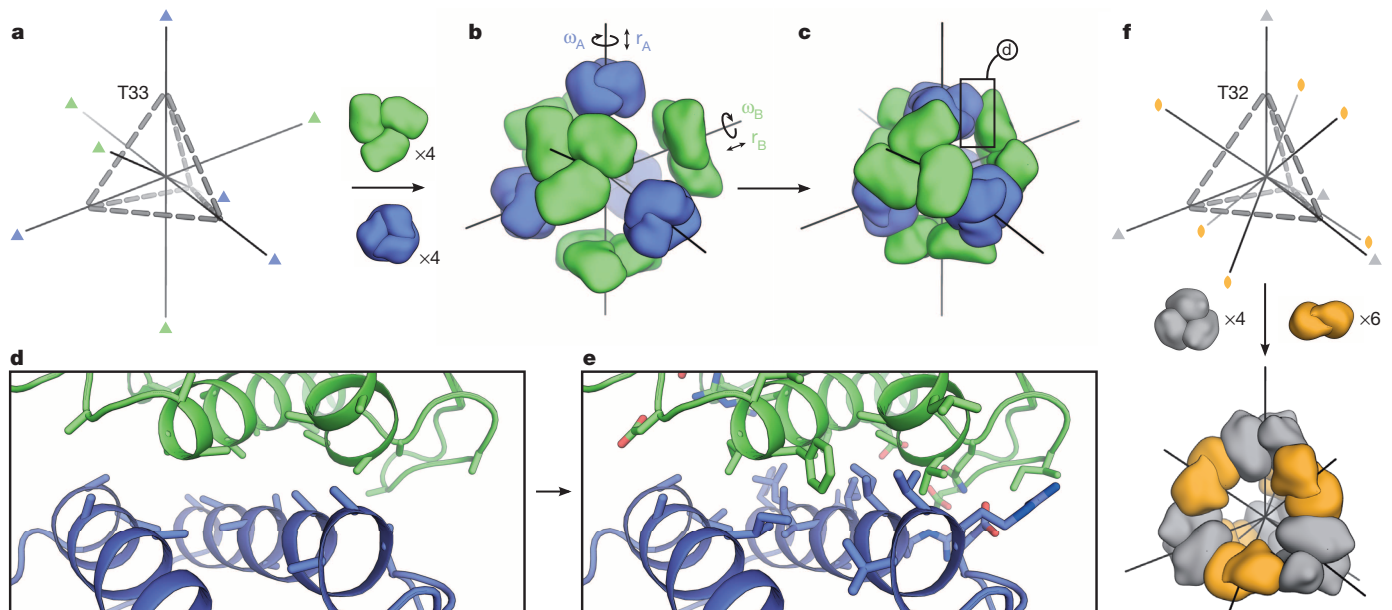
## Computational design method

Our method centres on encoding the information necessary to direct assembly in designed protein–protein interfaces. In addition to providing the energetic driving force for assembly, the designed interfaces also precisely define the relative orientations of the building blocks. We illustrate the method in Fig. 1 using the dual tetrahedral architecture (designated here as T33) as an example. In this architecture, four copies each of two distinct, naturally trimeric building blocks are aligned at opposite poles of the three-fold symmetry axes of a tetrahedron (Fig. 1a). This places one set of building blocks at the vertices of the tetrahedron and the other at the centres of the faces, totalling 12 subunits of each protein. Each trimeric building block is allowed to rotate around and translate along its three-fold symmetry axis (Fig. 1b); other rigid body moves are disallowed because they would lead to asymmetry. These four degrees of freedom are systematically explored during docking to identify configurations with symmetrically repeated instances of a novel inter-building-block interface that is suitable for design (Fig. 1c). The score function used during docking favours interfaces with high densities of

<sup>1</sup>Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. <sup>2</sup>Institute for Protein Design, University of Washington, Seattle, Washington 98195, USA. <sup>3</sup>Graduate Program in Molecular and Cellular Biology, University of Washington, Seattle, Washington 98195, USA. <sup>4</sup>UCLA Department of Chemistry and Biochemistry, Los Angeles, California 90095, USA. <sup>5</sup>Janelia Farm Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, Virginia 20147, USA. <sup>6</sup>UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, California 90095, USA. <sup>7</sup>UCLA Molecular Biology Institute, Los Angeles, California 90095, USA. <sup>8</sup>Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA.

\*These authors contributed equally to this work.



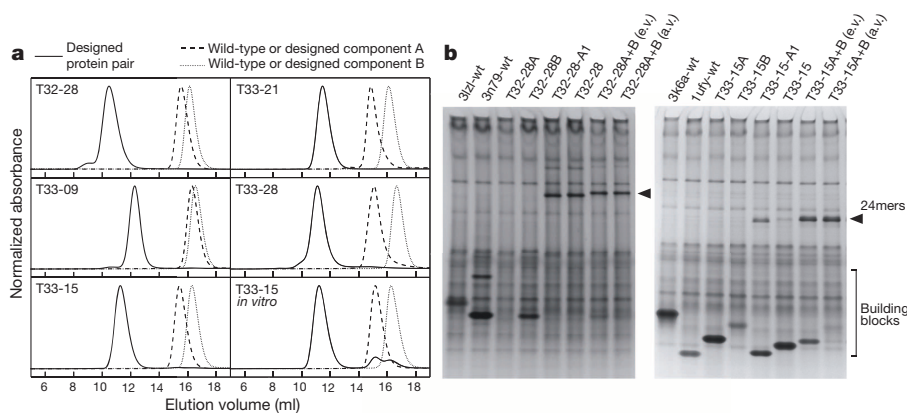


**Figure 1 | Overview of the computational design method.** **a**, The T33 architecture comprises four copies each of two distinct trimeric building blocks (green and blue) arranged with tetrahedral point group symmetry (24 total subunits; triangles indicate three-fold symmetry axes). **b**, Each building block has two rigid-body degrees of freedom, one translational ( $r$ ) and one rotational ( $\omega$ ), that are systematically explored during docking. **c**, The docking procedure, which is independent of the amino acid sequence of the building blocks, identifies large interfaces with high densities of contacting residues formed by

contacting residues in well-anchored regions of the protein structure that are less likely to change conformation on mutation of surface side chains (Fig. 1d). RosettaDesign<sup>30,31</sup> is then used to sample the identities and configurations of the side chains near the inter-building-block interface, generating interfaces with features resembling those found in natural protein assemblies such as well-packed hydrophobic cores surrounded by polar rims<sup>24</sup> (Fig. 1e). The end result is a pair of new amino acid sequences, one for each building block, predicted to stabilize the modelled interface and drive assembly to the specific target configuration.

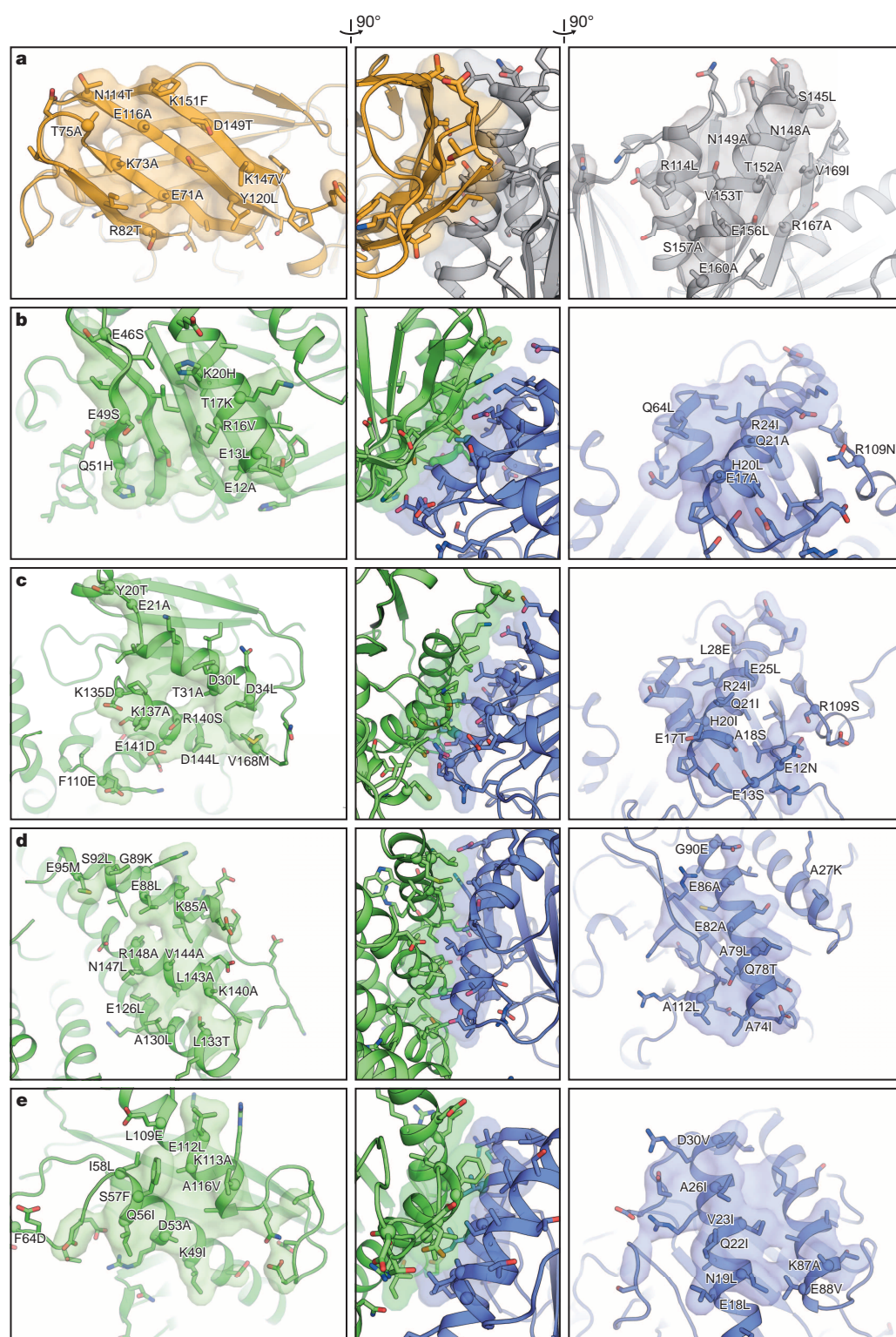
well-anchored regions of the protein structure. The details of such an interface, boxed here, are shown in **d**, **e**. Amino acid sequences are designed at the new interface to stabilize the modelled configuration and drive co-assembly of the two components. **f**, In the T32 architecture, four trimeric (grey) and six dimeric (orange) building blocks are aligned along the three-fold and two-fold symmetry axes passing through the vertices and edges of a tetrahedron, respectively.

These docking and design procedures were implemented by extending the Rosetta software<sup>31,32</sup> to enable the simultaneous modelling of multiple distinct symmetrically arranged protein components. The new protocol allows the different components to be arranged and moved independently according to distinct sets of symmetry operators (Extended Data Fig. 1). This enables the design strategy described above to be generalized to a wide variety of symmetric architectures in which multiple symmetric building blocks are combined in geometrically specific ways<sup>3,4,21</sup>. Combining even two types of symmetry elements (as in the present study) can give rise to a large number of distinct symmetric architectures with



**Figure 2 | Experimental characterization of co-assembly.** **a**, SEC chromatograms of the designed pairs of proteins (solid lines) and the wild-type oligomeric proteins from which they were derived (dashed and dotted lines). The co-expressed designed proteins elute at the volumes expected for the target 24-subunit nanomaterials, whereas the wild-type proteins elute as dimers or trimers. The T33-15 *in vitro* panel shows chromatograms for the individually produced and purified designed components (T33-15A and T33-15B) as well as a stoichiometric mixture of the two components. **b**, Native PAGE analysis of *in vitro*-assembled T32-28 (left panel) and T33-15 (right panel) in cell lysates.

Lysates containing the co-expressed design components (A1-tagged, lane 5; hexahistidine-tagged, lane 6) reveal slowly migrating species ('24mers', arrows) not present in lysates containing the wild-type or individually expressed components (lanes 1–4). Mixing equal volumes (e.v.) of crude lysates containing the individual designed components yields the same slowly migrating species (lane 7), although some unassembled building blocks remain due to unequal levels of expression (particularly for T33-15). When the differences in expression levels are accounted for by mixing adjusted volumes of lysates (a.v.), more efficient assembly is observed (lane 8).



**Figure 3 | Designed interfaces of two-component protein nanomaterials.** The models of the designed interfaces in each component (left, 'A' component; right, 'B' component) of T32-28 (a), T33-09 (b), T33-15 (c), T33-21 (d) and T33-28 (e) are shown, and side views of each interface as a whole are shown at centre (see arrows indicating rotations at top). Each image is oriented such that a vector originating at the centre of the tetrahedral material and passing through the centre of mass of the designed interface would pass vertically through the centre of the image. The side chains of all amino acids allowed to change identity or conformation during the interface design procedure are shown in stick representation. The alpha carbon atoms of positions that were mutated during design are shown as spheres, and the mutations are labelled. To highlight the morphologies of the contacting surfaces, atoms within 5 Å of the opposite building block are shown in semi-transparent surface representation. Oxygen atoms are red; nitrogen, blue; and sulphur, orange.

a range of possible morphologies, including those with dihedral and cubic point-group symmetries, as well as helical, layer and space group symmetries (ref. 21 and T.O.Y., manuscript in preparation).

In this study we targeted two distinct tetrahedral architectures: the T33 architecture described above and the T32 architecture shown in Fig. 1f, in which the materials are formed from four trimeric and six dimeric building blocks aligned along the three-fold and two-fold tetrahedral symmetry axes. We docked all pairwise combinations of a set of 1,161 dimeric and 200 trimeric protein building blocks of known structure in the T32 and T33 architectures (Supplementary Methods).

This resulted in a large set of potential novel nanomaterials: 232,200 and 19,900 docked protein pairs, respectively, with a given pair often yielding several distinct promising docked configurations. Interface sequence design calculations were carried out on the 1,000 highest scoring docked configurations in each architecture, and the designs were evaluated on the basis of predicted binding energy, shape complementarity<sup>33</sup> and size of the designed interfaces, as well as the number of buried unsatisfied hydrogen-bonding groups (Supplementary Methods). After filtering on these criteria, 30 T32 and 27 T33 materials were selected for experimental characterization (Extended Data Fig. 2). The 57 designs were



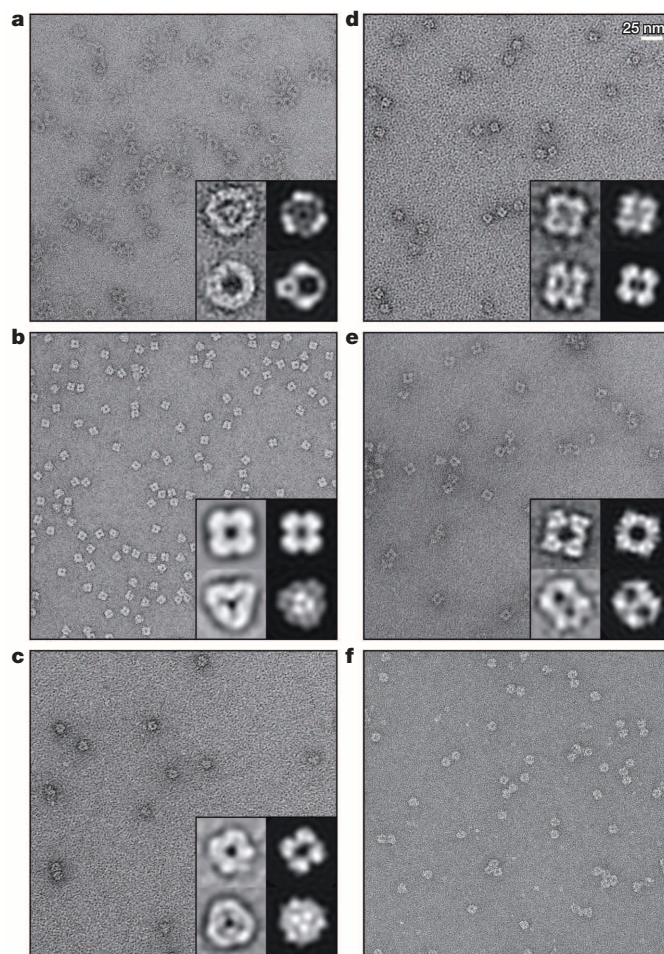
derived from 39 distinct trimeric and 19 dimeric proteins, and contained an average of 19 amino acid mutations per pair of subunits compared to the native sequences. The designed interfaces resided mostly on elements of secondary structure, both  $\alpha$ -helices and  $\beta$ -strands, with nearby loops generally making minor contributions.

### Screening and characterization of assembly state

Synthetic genes encoding each designed pair of proteins were cloned in tandem in a single expression vector to allow inducible co-expression in *Escherichia coli* (Supplementary Methods). Polyacrylamide gel electrophoresis (PAGE) under denaturing and non-denaturing (native) conditions was used to rapidly assess the level of soluble expression and assembly state of the designed proteins in clarified cell lysates. For most of the designs, either one or both of the designed proteins was not detectable in the soluble fraction, suggesting that insoluble expression is a common failure mode for the designed materials. Given that the majority of the mutations introduced by our method are changes from polar to hydrophobic residues at the designed interfaces, it is likely that the insolubility of these designs is due to either misfolding or non-specific aggregation of the designed protein subunits. Nevertheless, several designed protein pairs yielded single bands under non-denaturing conditions that migrated more slowly than the wild-type proteins from which they were derived, suggesting assembly to higher-order species (Extended Data Fig. 3). These proteins were subcloned to introduce a hexahistidine tag at the carboxy terminus of one of the two subunits and purified by nickel affinity chromatography and size exclusion chromatography (SEC). Five pairs of designed proteins—one T32 design (T32-28) and four T33 designs (T33-09, T33-15, T33-21 and T33-28)—eluted together during nickel affinity chromatography and yielded dominant peaks at the expected size of approximately 24 subunits when analysed by SEC (Fig. 2a and Supplementary Table 1).

We tested the ability of each of the five materials to assemble *in vitro* by expressing the two components in separate *E. coli* cultures and mixing them at various points after cell lysis (Extended Data Fig. 3). Native PAGE revealed that in two cases (T33-15 and T32-28) the two separately expressed components efficiently assembled to give the designed materials *in vitro* when equal volumes of cell lysates were mixed (Fig. 2b, Extended Data Fig. 3a, c). Adjusting the volume of each lysate in the mixture to account for differences in the level of soluble expression of the two components allowed for more quantitative assembly. In the case of T33-15, the two components of the material could also be purified independently: T33-15A and T33-15B each eluted from the SEC column as trimers in isolation. After mixing the two purified components in a 1:1 molar ratio and allowing a two-hour incubation at room temperature, the mixture eluted from the SEC column predominantly at the volume expected for the 24-subunit assembly, with small amounts of residual trimeric building blocks remaining (Fig. 2a). It is thus possible to control the assembly of our designed materials by simply mixing the two independently produced components.

The details of the designed interfaces for the five materials, highlighting the shape and chemical complementarity generated by the many amino acid mutations introduced during design, are presented in Fig. 3. Qualitatively, the interfaces reflect the hypothesis underlying the design protocol: they feature well-packed and highly complementary cores of hydrophobic side chains residing mostly in elements of secondary structure, surrounded by polar side chains lining the periphery of the hydrophobic cores. The successful designs are quantitatively similar to the other designs according to the interface metrics used to select designs for experimental characterization (predicted binding energy, shape complementarity, interface size and number of buried unsatisfied hydrogen-bonding groups; Extended Data Fig. 4). The similarity of the successful and unsuccessful designs according to these structural metrics, combined with the observed insolubility of many of the designs, suggests that focusing on improving the level of soluble expression of the designed proteins could substantially improve the success rate of our approach in the future.



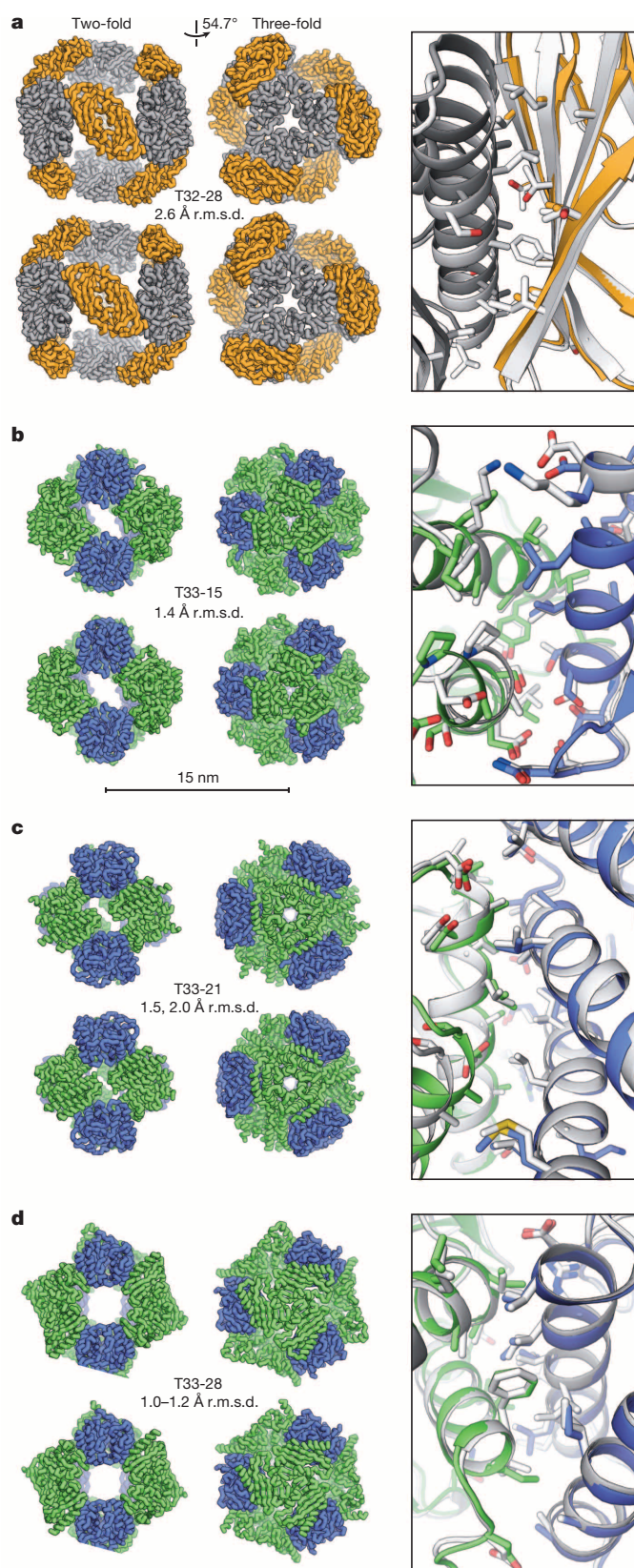
**Figure 4 | Electron micrographs of designed two-component protein nanomaterials.** Negative-stain electron micrographs for co-expressed and purified T32-28 (a), T33-09 (b), T33-15 (c), T33-21 (d) and T33-28 (e) are shown to scale (scale bar at top right, 25 nm). For each co-expressed material, two different class averages of the particles (top and bottom) are shown in the insets (left) alongside back projections calculated from the computational design models (right). f, Micrograph of a T33-15 sample prepared by stoichiometrically mixing the independently purified components (T33-15A and T33-15B) *in vitro* and purifying the assembled material by SEC (see Fig. 2). Micrographs of unpurified, *in vitro*-assembled T33-15 as well as T33-15A and T33-15B in isolation are shown in Extended Data Fig. 5.

### Structural characterization of the designed materials

Negative-stain electron microscopy of the five designed materials confirmed that they assemble specifically to the target architectures (Fig. 4). For each material, fields of monodisperse particles of the expected size and symmetry were observed, confirming the homogeneity of the materials suggested by SEC. Particle averaging yielded images that recapitulate features of the computational design models at low resolution. For example, class averages of T33-09 revealed roughly square or triangle-shaped structures with well-defined internal cavities that closely resemble projections calculated from the computational design model along its two-fold and three-fold axes (Fig. 4b, inset). Micrographs of T33-15 assembled *in vitro* as described above were indistinguishable from those of co-expressed T33-15 (Fig. 4c, f and Extended Data Fig. 5c), demonstrating that the same material is obtained using both methods.

We solved X-ray crystal structures of four of the designed materials (T32-28, T33-15, T33-21 and T33-28) to resolutions ranging from 2.1 to 4.5 Å (Fig. 5 and Supplementary Tables 2 and 3). In all cases, the structures reveal that the inter-building-block interfaces were designed with high accuracy: comparing a pair of chains from each structure to the computationally designed model yields backbone root mean square





deviations (r.m.s.d.) between 0.5 and 1.2 Å (Fig. 5 right and Extended Data Table 1). In the structures with resolutions that permit detailed analysis of side-chain configurations (T33-15 and two independent crystal forms of T33-21), 87 of 113 side chains at the designed interfaces adopt the predicted conformations (Supplementary Tables 5 and 6). As intended, the designed interfaces drive the assembly of cage-like

**Figure 5 | Crystal structures of designed two-component protein nanomaterials.** The computational design models (top) and X-ray crystal structures (bottom) are shown at left for T32-28 (**a**), T33-15 (**b**), T33-21 (**c**) and T33-28 (**d**). Views of each material are shown to scale along the two-fold and three-fold tetrahedral symmetry axes (scale bar at centre, 15 nm). The r.m.s.d. values between the backbone atoms in all 24 chains of the design models and crystal structures are indicated. For T33-21 (**c**), r.m.s.d. values are shown for both crystal forms (images are shown for the higher-resolution crystal form with backbone r.m.s.d. 2.0 Å), while the r.m.s.d. range for T33-28 (**d**) derives from the four copies of the fully assembled material in the crystallographic asymmetric unit. At right, overlays of the designed interfaces in the design models (white) and crystal structures (grey, orange, green and blue) are shown. Owing to the limited resolution of the T32-28 structure, the amino acid side chains were not modelled beyond the beta carbon. For the interface overlays, the crystal structures were aligned to the design models using the backbone atoms of two subunits, one of each component.

nanomaterials that closely match the computational design models: the backbone r.m.s.d. over all 24 subunits in each material range from 1.0 to 2.6 Å (Fig. 5 left and Extended Data Table 1). The precise control over interface geometry offered by our method thus enables the design of two-component protein nanomaterials with diverse nanoscale features such as surfaces, pores and internal volumes with high accuracy.

## Discussion

Owing to the unique functions accessible to self-assembling proteins, there is intense interest in engineering protein nanomaterials for applications in various fields. Most efforts so far have focused on repurposing naturally occurring protein assemblies, a strategy that is ultimately limited by the structures available and their tolerances for modification. Similarly, although directed evolution is a powerful method for protein engineering<sup>34,35</sup> and can be used to improve, for example, the packaging capability of existing protein nanocontainers<sup>36,37</sup>, it is difficult to envision how it could accurately generate new protein nanomaterials with target structures defined at the atomic level. Our results demonstrate that computational protein design provides a general route for designing novel two-component self-assembling protein nanomaterials with high accuracy. The combinatorial nature of two-component materials greatly expands the number and variety of potential nanomaterials that can be designed. For example, in this study we used 1,361 protein building blocks to dock over 250,000 distinct protein pairs in two target architectures with tetrahedral point group symmetry, resulting in a very large set of potential nanomaterials exhibiting a variety of sizes, shapes and arrangements of chemically and genetically addressable functional groups, loops and termini. With continued effort to increase the success rate of protein–protein interface design and reduce the rate of designed proteins that express insolubly, it should become possible to simultaneously design multiple novel interfaces in a single material, which would enable the construction of increasingly complex materials built from more than two components.

The conceptual framework that underlies our method—symmetric docking followed by protein–protein interface design—can be generally applied to a wide variety of symmetric architectures, including repetitive protein arrays that extend in one, two or three dimensions. Multi-component materials are advantageous in these extended architectures because the uncontrolled self-assembly of a single-component material inside the cell can complicate biological production<sup>5,11,21</sup>. We have shown that the two components of the designed materials T32-28 and T33-15 can be produced separately and mixed *in vitro* to initiate assembly of the designed structure. With new symmetric modelling algorithms capable of handling the additional degrees of freedom associated with these architectures, the accurate computational design and controllable assembly of complex, multi-component protein fibres, layers and crystals should also be possible.

The capability to design highly homogeneous protein nanostructures with atomic-level accuracy and controllable assembly should open up new opportunities in targeted drug delivery, vaccine design, plasmonics

and other applications that can benefit from the precise patterning of matter on the subnanometre to 100-nanometre scale. Extending beyond static structure design, methods for incorporating the kinds of dynamic and functional behaviours observed in natural protein assemblies should make possible the design of novel protein-based molecular machines with programmable structures, dynamics and functions.

## METHODS SUMMARY

The symmetric modelling framework in Rosetta<sup>31,32</sup> was updated to enable the modelling of multi-component symmetrical structures. A new application, *tcdock*, docks pairs of protein scaffolds in higher-order symmetries, scoring each docked configuration according to its suitability for interface design. *tcdock* was used to dock all possible pairwise combinations of 200 trimeric scaffold proteins and all possible pairwise combinations of the same trimers and 1,161 dimeric proteins in the T33 and T32 symmetric architectures, respectively. New two-component protein–protein interface design protocols were used to design new amino acid sequences predicted to stabilize selected docked configurations. During the sequence design protocols, the symmetric rigid body degrees of freedom and the identities and conformations of the side chains at the inter-building-block interfaces were optimized to identify low-energy sequence–structure combinations. Thirty T32 and 27 T33 designs were selected for experimental characterization.

The assembly states of the designed pairs of proteins were assessed by native PAGE, and those that migrated more slowly than the wild-type scaffolds were subjected to affinity purification and SEC. The ability of the materials to assemble *in vitro* was investigated by independently producing the two components, mixing them at various points after cell lysis, and analysing the mixtures by native PAGE and SEC. The materials were structurally characterized by negative-stain electron microscopy and particle averaging, and at high resolution by X-ray crystallography.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 January; accepted 25 April 2014.

Published online 25 May 2014.

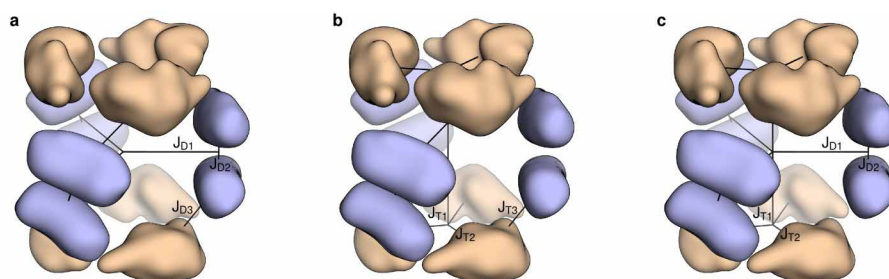
- Howorka, S. Rationally engineering natural protein assemblies in nanobiotechnology. *Curr. Opin. Biotechnol.* **22**, 485–491 (2011).
- Douglas, T. & Young, M. Viruses: making friends with old foes. *Science* **312**, 873–875 (2006).
- Lai, Y. T., King, N. P. & Yeates, T. O. Principles for designing ordered protein assemblies. *Trends Cell Biol.* **22**, 653–661 (2012).
- King, N. P. & Lai, Y. T. Practical approaches to designing novel protein assemblies. *Curr. Opin. Struct. Biol.* **23**, 632–638 (2013).
- Sinclair, J. C. Constructing arrays of proteins. *Curr. Opin. Chem. Biol.* **17**, 946–951 (2013).
- Salgado, E. N., Radford, R. J. & Tezcan, F. A. Metal-directed protein self-assembly. *Acc. Chem. Res.* **43**, 661–672 (2010).
- King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174 (2012).
- Brodin, J. D. *et al.* Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nature Chem.* **4**, 375–382 (2012).
- Lanci, C. J. *et al.* Computational design of a protein crystal. *Proc. Natl Acad. Sci. USA* **109**, 7304–7309 (2012).
- Stranges, P. B., Machius, M., Miley, M. J., Tripathy, A. & Kuhlman, B. Computational design of a symmetric homodimer using beta-strand assembly. *Proc. Natl Acad. Sci. USA* **108**, 20562–20567 (2011).
- Sinclair, J. C., Davies, K. M., Venien-Bryan, C. & Noble, M. E. Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nature Nanotechnol.* **6**, 558–562 (2011).
- Lai, Y. T., Cascio, D. & Yeates, T. O. Structure of a 16-nm cage designed by using protein oligomers. *Science* **336**, 1129 (2012).
- Der, B. S. *et al.* Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J. Am. Chem. Soc.* **134**, 375–385 (2012).
- Fletcher, J. M. *et al.* Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595–599 (2013).
- Boyle, A. L. *et al.* Squaring the circle in peptide assembly: from fibers to discrete nanostructures by de novo design. *J. Am. Chem. Soc.* **134**, 15457–15467 (2012).
- Grigoryan, G. *et al.* Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* **332**, 1071–1076 (2011).
- Seeman, N. C. Nanomaterials based on DNA. *Annu. Rev. Biochem.* **79**, 65–87 (2010).
- Rothmund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
- Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-dimensional structures self-assembled from DNA bricks. *Science* **338**, 1177–1183 (2012).
- Han, D. *et al.* DNA gridiron nanostructures based on four-arm junctions. *Science* **339**, 1412–1415 (2013).
- Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl Acad. Sci. USA* **98**, 2217–2221 (2001).
- Usui, K. *et al.* Nanoscale elongating control of the self-assembled protein filament with the cysteine-introduced building blocks. *Protein Sci.* **18**, 960–969 (2009).
- Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
- Janin, J., Bahadur, R. P. & Chakrabarti, P. Protein–protein interaction and quaternary structure. *Q. Rev. Biophys.* **41**, 133–180 (2008).
- Huang, P. S., Love, J. J. & Mayo, S. L. A de novo designed protein protein interface. *Protein Sci.* **16**, 2770–2774 (2007).
- Jha, R. K. *et al.* Computational design of a PAK1 binding protein. *J. Mol. Biol.* **400**, 257–270 (2010).
- Karanicolas, J. *et al.* A de novo protein binding pair by computational design and directed evolution. *Mol. Cell* **42**, 250–260 (2011).
- Fleishman, S. J. *et al.* Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816–821 (2011).
- Khare, S. D. & Fleishman, S. J. Emerging themes in the computational design of novel enzymes and protein–protein interfaces. *FEBS Lett.* **587**, 1147–1154 (2013).
- Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl Acad. Sci. USA* **97**, 10383–10388 (2000).
- Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
- DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D. & Andre, I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS ONE* **6**, e20450 (2011).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Arnold, F. H. & Volkov, A. A. Directed evolution of biocatalysts. *Curr. Opin. Chem. Biol.* **3**, 54–59 (1999).
- Jäckel, C., Kast, P. & Hilvert, D. Protein design by directed evolution. *Annu. Rev. Biophys.* **37**, 153–173 (2008).
- Wörsdörfer, B., Pianowski, Z. & Hilvert, D. Efficient *in vitro* encapsulation of protein cargo by an engineered protein container. *J. Am. Chem. Soc.* **134**, 909–911 (2012).
- Wörsdörfer, B., Woycechowsky, K. J. & Hilvert, D. Directed evolution of a protein container. *Science* **331**, 589–592 (2011).
- Bradley, P. & Baker, D. Improved beta-protein structure prediction by multilevel optimization of nonlocal strand pairings and local backbone conformation. *Proteins* **65**, 922–929 (2006).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank D. Shi and B. Nannenga (JFRC) for help with electron microscopy, F. DiMaio and R. Moretti for assistance with software development, P. Greisen for scripts used to compare side-chain conformations, J. Gallaher for technical assistance, M. Collazo for help with preliminary crystallization screening, D. Cascio and M. Sawaya for help with crystallographic experiments, and M. Capel, J. Schuermann and I. Kourinov at NE-CAT beamline 24-ID-C for help with data collection. This work was supported by the Howard Hughes Medical Institute (T.G. and D.B.) and the JFRC visitor program (S.G.), the National Science Foundation under CHE-1332907 (D.B. and T.O.Y.), grants from the International AIDS Vaccine Initiative, DTRA (N00024-10-D-6318/0024), AFOSR (FA950-12-1-0112) and DOE (DE-SC0005155) to D.B., an NIH Biotechnology Training Program award to D.E.M. (T32GM067555) and an NSF graduate research fellowship to J.B.B. (DGE-0718124). T.O.Y. and D.E.M. also acknowledge support from the BER programme of the DOE Office of Science. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding bodies.

**Author Contributions** N.P.K., J.B.B., W.S. and D.B. designed the research. N.P.K., J.B.B. and W.S. wrote program code and performed the docking and design calculations. N.P.K. and J.B.B. biophysically characterized the designed materials and prepared samples for structural analysis. S.G. characterized the designed materials by electron microscopy; S.G. and T.G. analysed electron microscopy data. D.E.M. crystallized the designed protein materials; D.E.M. and T.O.Y. analysed crystallographic data. N.P.K., J.B.B. and D.B. analysed data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

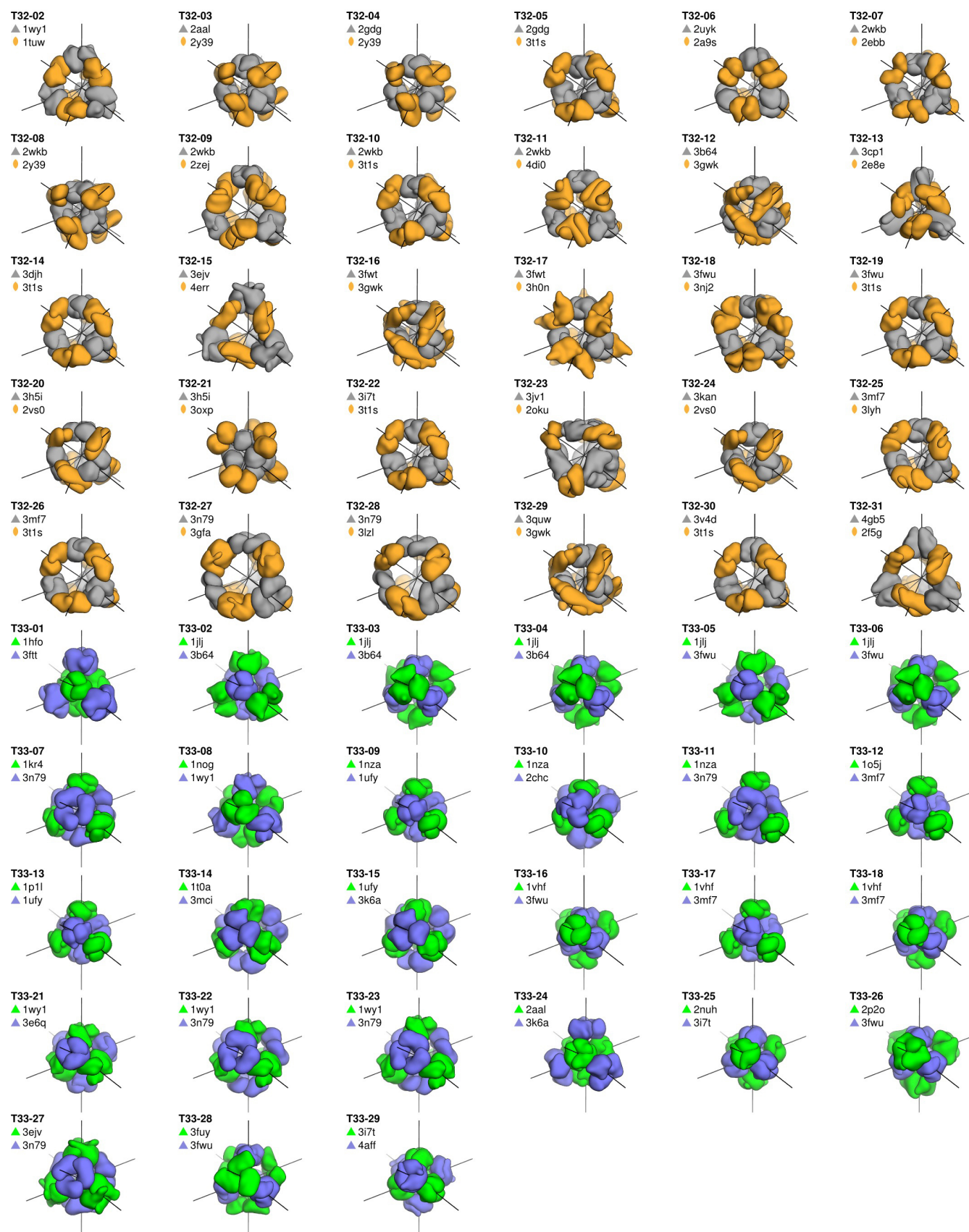
**Author Information** The crystal structures and structure factors for the designed materials have been deposited in the RCSB Protein Data Bank (<http://www.rcsb.org/>) under the accession codes 4NWN (T32-28), 4NWO (T33-15), 4NWP (T33-21, R32 crystal form), 4NWQ (T33-21, F4; 32 crystal form) and 4NWR (T33-28). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B. ([dabaker@u.washington.edu](mailto:dabaker@u.washington.edu)).



**Extended Data Figure 1 | Comparison of one-component and multi-component symmetric fold trees.** Within the Rosetta macromolecular modelling suite, the connections between residues in a protein structure are represented as a directed, acyclic, graph referred to as a ‘fold tree’<sup>32,38</sup>. When modelling multiple subunits in symmetric systems, the rigid body orientations of the subunits can be controlled by modifying the appropriate connections in the fold tree. In this work, we have extended Rosetta to allow multiple, independently managed connections from the fold tree to the subunits in the asymmetric unit (ASU) of the modelled structure. To demonstrate the new behaviour enabled by this change, three different symmetric fold tree representations of a D32 architecture are shown. In this architecture, which is used because of its relative simplicity, two trimeric building blocks (wheat) are aligned along the three-fold rotational axes of D3 point-group symmetry and three dimeric building blocks (light blue) are aligned along the two-fold rotational axes. **a**, The dimer-centric one-component symmetry case.

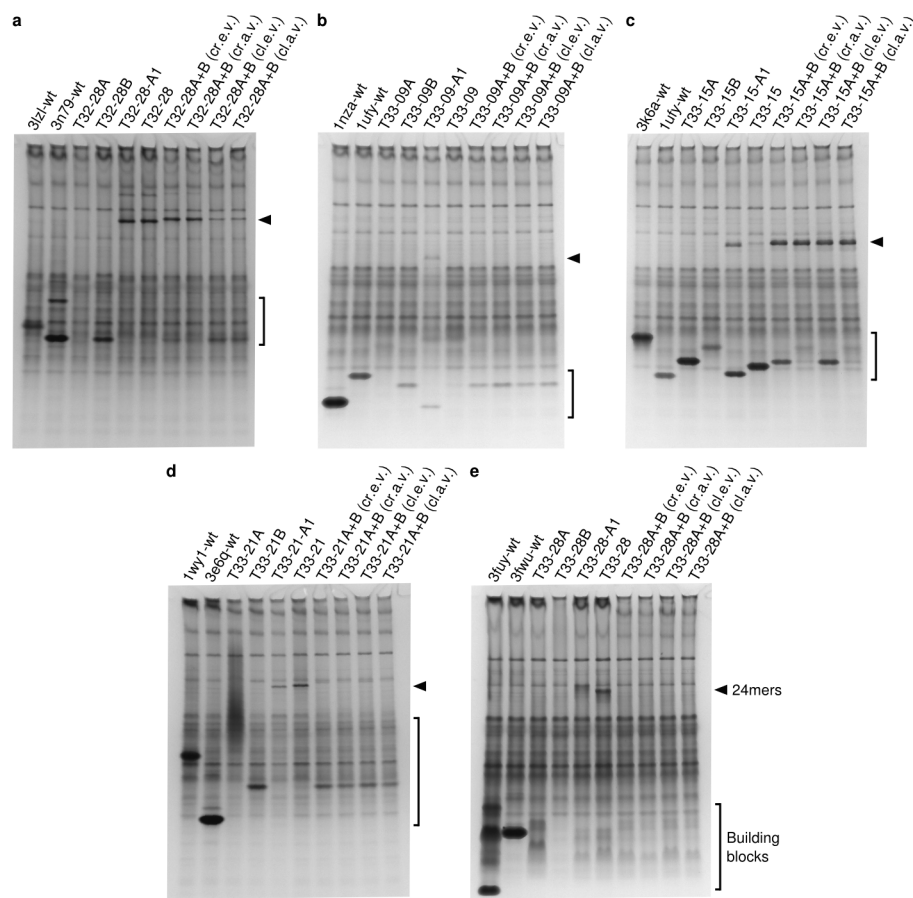
Rigid-body degree of freedom (RB DOF, black lines)  $J_{D3}$  connecting the dimer subunit to the trimer subunit in the ASU is downstream of RB DOFs  $J_{D1}$  and  $J_{D2}$  controlling the dimer subunit; in this case the positions of the trimeric subunits depend on the positions of the dimeric subunits. **b**, The trimer-centric one-component symmetry case. RB DOF  $J_{T3}$  connecting the trimer subunit to the dimer subunit in the ASU is downstream of RB DOFs  $J_{T1}$  and  $J_{T2}$  controlling the trimer subunit; in this case the positions of the dimeric subunits depend on the positions of the trimeric subunits. **c**, The multi-component symmetry case. With multi-component symmetric modelling, the RB DOFs controlling the trimer subunit ( $J_{T1}$  and  $J_{T2}$ ) and the dimer subunit ( $J_{D1}$  and  $J_{D2}$ ) in the ASU are independent. In this case the positions of the dimeric subunits do not depend on the positions of the trimeric subunits and vice versa, allowing the internal DOFs for each building block ( $J_{T2}$  and  $J_{D2}$ ) to be maintained while moving the building blocks independently ( $J_{T1}$  and  $J_{D1}$ ). See Supplementary Methods for additional discussion.





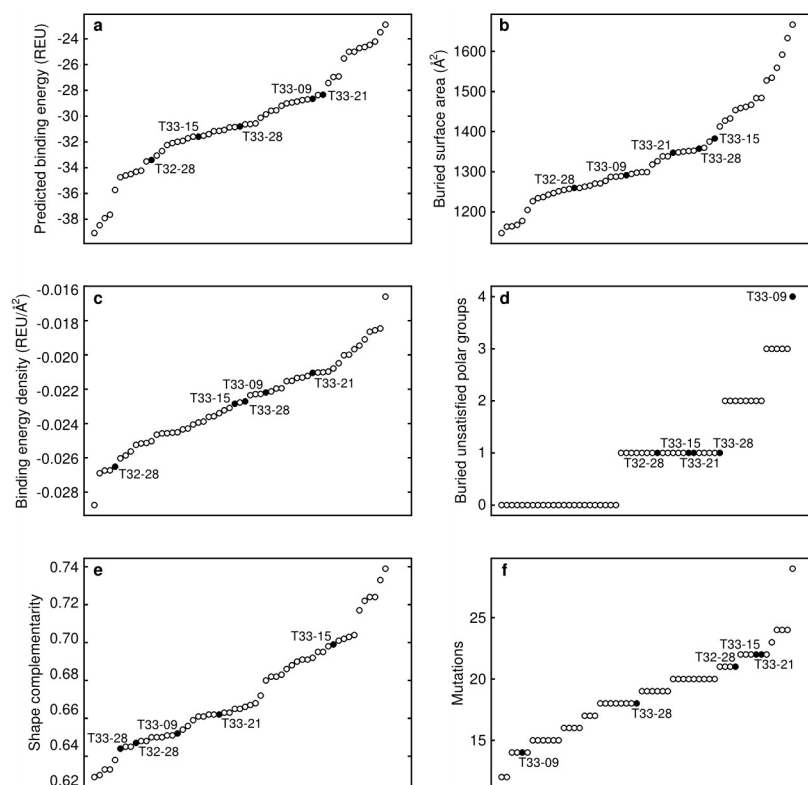
**Extended Data Figure 2 | Models of the 57 designs selected for experimental characterization.** Smoothed surface representations of each of the 30 T32 and 27 T33 designs are shown. The trimeric component of each T32 design is shown in grey and the dimeric component in orange. The two different trimeric components of each T33 design are shown in blue and green. The tetrahedral

two-fold and three-fold symmetry axes (black lines) are shown passing through the centre of each component. Each design is named according to its symmetric architecture (T32 or T33) followed by a unique identification number. The pairs of scaffold proteins from which the designs are derived are indicated.



**Extended Data Figure 3 | Native PAGE analysis of cleared cell lysates.** Each gel contains cleared lysates pertaining to **a**, T32-28, **b**, T33-09, **c**, T33-15, **d**, T33-21, or **e**, T33-28. Lane 1 is from cells expressing the wild-type scaffold for component A and lane 2 the wild-type scaffold for component B. Lanes 3 and 4 are from cells expressing the individual design components and lanes 5 and 6 the co-expressed components. Lanes 7 and 8 are from samples mixed as crude equal volume or crude adjusted volume (cr.e.v. or cr.a.v.) lysates, while

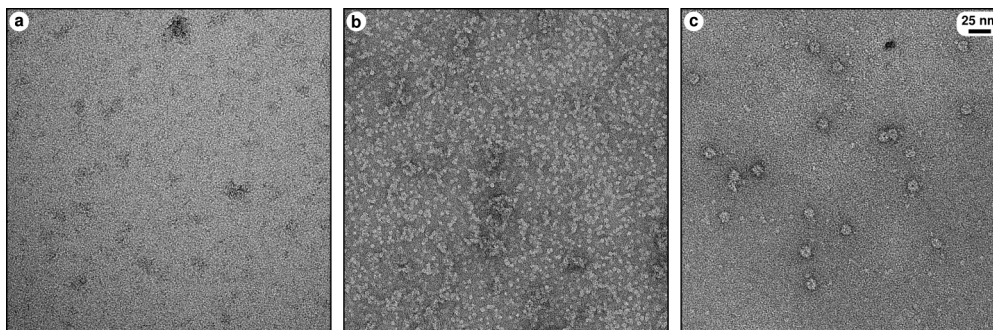
lanes 9 and 10 are from samples mixed as cleared lysates (cl.e.v. or cl.a.v.). Lane 5 is from cells expressing the C-terminally A1-tagged constructs; all other lanes are from cells expressing the C-terminally His-tagged constructs. An arrow is positioned next to each gel indicating the migration of 24-subunit assemblies and the gel regions containing unassembled building blocks are bracketed. Each gel was stained with GelCode Blue. Portions of the gels in **a** and **c** are also shown in Fig. 2b.



**Extended Data Figure 4 | Structural metrics for the computational design models.** Selected metrics related to the designed interfaces are plotted for the 57 designs that were experimentally characterized, including **a**, the predicted binding energy measured in Rosetta energy units (REU), **b**, the surface area buried by each instance of the designed interface, **c**, the binding energy density (calculated as the predicted binding energy divided by the buried surface area),

**d**, the number of buried unsatisfied polar groups at the designed interface, **e**, the shape complementarity of the designed interface, and **f**, the total number of mutations in each designed pair of proteins. Each circle represents a single design; the five successful materials are plotted as filled circles and labelled. In each plot, the designs are arranged on the *x* axis in order of increasing value of the metric analysed.





**Extended Data Figure 5 | Electron micrographs of *in vitro*-assembled T33-15.** Negative stain micrographs of independently purified T33-15A (a) and T33-15B (b), as well as unpurified, *in vitro*-assembled T33-15 (c) are shown to scale (scale bar at right, 25 nm).

**Extended Data Table 1 | Root mean square deviations (r.m.s.d.) between crystal structures and design models**

Design model	Crystal structure	Global r.m.s.d. (Å)*	Two-chain r.m.s.d. (Å)†	Contents of asymmetric unit	Structure used for superposition‡
T32-28	4NWN	2.586	1.246	One cage (24 subunits)	Asymmetric unit
T33-15	4NWO	1.433	0.876	One chain of each component (2 subunits)	One cage generated from crystallographic 2- and 3-folds
T33-21	4NWP	1.962	0.924	4 chains of each component (8 subunits)	One cage generated from one crystallographic 3-fold
T33-21	4NWO	1.482	0.765	One chain of each component (2 subunits)	One cage generated from crystallographic 2- and 3-folds
T33-28	4NWR	0.965	0.503	Four complete cages (96 subunits)	One complete cage from the asymmetric unit
T33-28	4NWR	0.965	0.548	Four complete cages (96 subunits)	One complete cage from the asymmetric unit
T33-28	4NWR	1.195	0.567	Four complete cages (96 subunits)	One complete cage from the asymmetric unit
T33-28	4NWR	1.212	0.477	Four complete cages (96 subunits)	One complete cage from the asymmetric unit

\*Global backbone r.m.s.d. was calculated over all 24 subunits of each design model and corresponding subunits in each crystal structure.

†Two-chain backbone r.m.s.d. was calculated over chains A and B of each design model and corresponding subunits in each crystal structure.

‡24 subunits composing one complete cage were derived from each crystal structure as indicated and the chains renamed to match the corresponding names in the design models. In the case of T33-28, four different sets of r.m.s.d. calculations were carried out, one for each of the four cages contained in the asymmetric unit of 4NWR.

# The ctenophore genome and the evolutionary origins of neural systems

Leonid L. Moroz<sup>1,2,3</sup>, Kevin M. Kocot<sup>4</sup>, Mathew R. Citarella<sup>1</sup>, Sohn Dosung<sup>1</sup>, Tigran P. Norekian<sup>1,3</sup>, Inna S. Povolotskaya<sup>5,6</sup>, Anastasia P. Grigorenko<sup>7,8</sup>, Christopher Dailey<sup>9</sup>, Eugene Berezikov<sup>10</sup>, Katherine M. Buckley<sup>11</sup>, Andrey Ptitsyn<sup>1</sup>, Denis Reshetov<sup>8</sup>, Krishanu Mukherjee<sup>1</sup>, Tatiana P. Moroz<sup>1</sup>, Yelena Bobkova<sup>1</sup>, Fahong Yu<sup>2</sup>, Vladimir V. Kapitonov<sup>12</sup>, Jerzy Jurka<sup>12</sup>, Yuri V. Bobkov<sup>1</sup>, Joshua J. Swore<sup>1,3</sup>, David O. Girardo<sup>1,3</sup>, Alexander Fodor<sup>1</sup>, Fedor Gusev<sup>7,8</sup>, Rachel Sanford<sup>1</sup>, Rebecca Bruders<sup>1,3</sup>, Ellen Kittler<sup>13</sup>, Claudia E. Mills<sup>3</sup>, Jonathan P. Rast<sup>11</sup>, Romain Derelle<sup>5,6</sup>, Victor V. Solovyev<sup>14</sup>, Fyodor A. Kondrashov<sup>5,6,15</sup>, Billie J. Swalla<sup>3</sup>, Jonathan V. Sweedler<sup>9</sup>, Evgeny I. Rogayev<sup>7,8,16,17</sup>, Kenneth M. Halanych<sup>4</sup> & Andrea B. Kohn<sup>1</sup>

**The origins of neural systems remain unresolved. In contrast to other basal metazoans, ctenophores (comb jellies) have both complex nervous and mesoderm-derived muscular systems. These holoplanktonic predators also have sophisticated ciliated locomotion, behaviour and distinct development. Here we present the draft genome of *Pleurobrachia bachei*, Pacific sea gooseberry, together with ten other ctenophore transcriptomes, and show that they are remarkably distinct from other animal genomes in their content of neurogenic, immune and developmental genes. Our integrative analyses place Ctenophora as the earliest lineage within Metazoa. This hypothesis is supported by comparative analysis of multiple gene families, including the apparent absence of HOX genes, canonical microRNA machinery, and reduced immune complement in ctenophores. Although two distinct nervous systems are well recognized in ctenophores, many bilaterian neuron-specific genes and genes of 'classical' neurotransmitter pathways either are absent or, if present, are not expressed in neurons. Our metabolomic and physiological data are consistent with the hypothesis that ctenophore neural systems, and possibly muscle specification, evolved independently from those in other animals.**

Approximately 150 recognized species of comb jellies form a clade of pre-bilaterian animals<sup>1–3</sup> (Fig. 1f) with an elusive genealogy, possibly tracing their ancestry to the Ediacaran biota<sup>4,5</sup>. We selected the Pacific sea gooseberry, *Pleurobrachia bachei* (A. Agassiz, 1860; Figs 1a, 2, Extended Data Fig. 1, Supplementary Data 1 and Supplementary Videos 1 and 2) as a model ctenophore because of preserved traits thought to be ancestral for this lineage (for example, cydippid larva and tentacles). Three next-generation sequencing platforms (454, Illumina and Ion Torrent) were used to obtain >700-fold coverage (Supplementary Tables 1 and 2) of *Pleurobrachia*'s genome, and about 2,000-fold coverage of the transcriptome representing all major organs and developmental stages (Supplementary Tables 3 and 4). Consequently, the draft assembly was 156,146,497 base pairs (bp) with 19,523 predicted protein-coding genes (Supplementary Tables 5–7). About 90% of these predicted genes are expressed in at least one tissue or developmental stage (Supplementary Methods) and 44% of *Pleurobrachia* genes have orthologues in other animals (Supplementary Tables 7 and 8). More than 300 families of transposable elements constitute at least 8.5% of the genome (Supplementary Table 9 and Supplementary Data 2) with numerous examples of diversification of some ancient transposable element classes (for example, transposases and reverse transcriptases). Approximately 1.0% of the genome is methylated. *Pleurobrachia* also uses DNA demethylation during development, with both 5-hydroxymethyl cytosine (5-hmC) and its synthetic enzyme

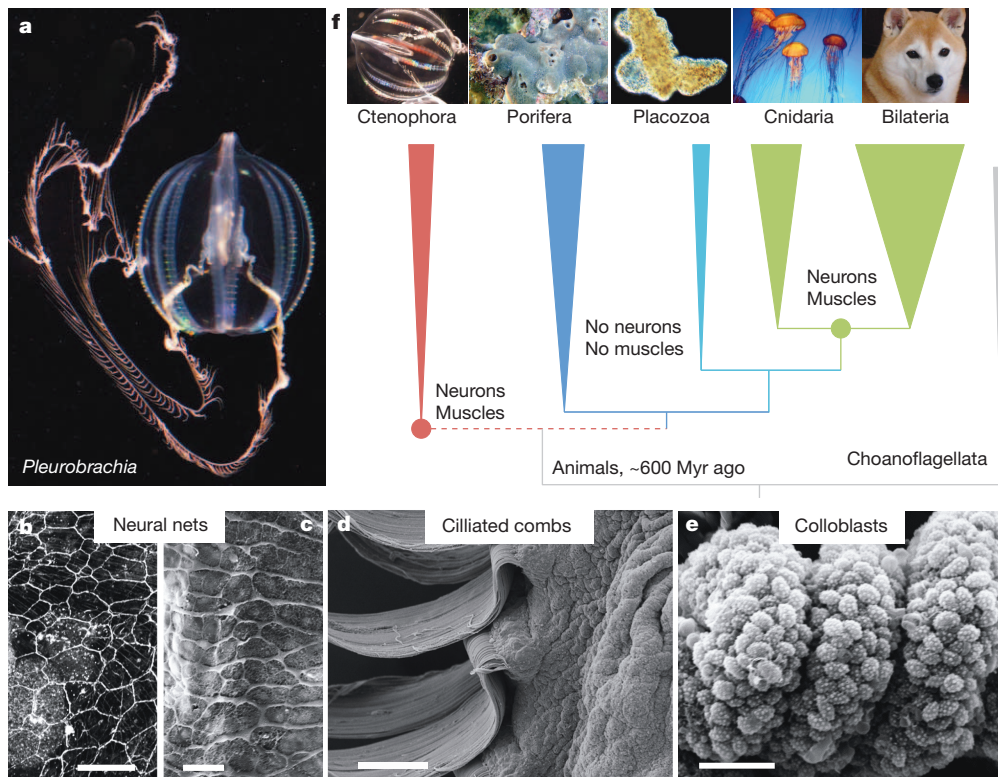
TET<sup>6</sup> (Extended Data Fig. 2). The obtained genome and transcriptome data provide rich resources (<http://neurobase.rc.ufl.edu/Pleurobrachia>) for investigating both animal phylogeny and evolution of animal innovations including nervous systems<sup>2,3,7–9</sup>.

## Ctenophore phylogeny

Although relationships among basal animal lineages are controversial<sup>11,10–16</sup>, our analyses (Supplementary Data 4) with Ctenophora represented by *Pleurobrachia* and *Mnemiopsis* suggest the placement of Ctenophora as the most basal animal lineage (Fig. 1 and Extended Data Fig. 3). Porifera was recovered as sister taxon to remaining metazoans (bootstrap support = 100%) with Cnidaria sister to Bilateria (bootstrap support = 100%, Fig. 1f). Shimodaira–Hasegawa tests<sup>17</sup> (corresponding to Extended Data Fig. 3a–c with 586 gene matrix) rejected both Eumetazoa (sponges sister group to all other metazoans) and Coelenterata (Cnidaria + Ctenophora). Placement of Ctenophora at the base of Metazoa also provides the most parsimonious explanation for the pattern of global gene gain/loss seen across major animal clades (Fig. 3a and Supplementary Table 14a, b). Transcriptome data from ten additional ctenophores (Supplementary Table 13) with stricter criteria for orthology inference (Supplementary Methods 7) also placed ctenophores basal, albeit with less support (Extended Data Fig. 3d). When the most conserved set of genes was considered (Supplementary Methods 7.5 and Supplementary Data 4.3), the

<sup>1</sup>The Whitney Laboratory for Marine Bioscience, University of Florida, 9505 Ocean Shore Blvd, St Augustine, Florida 32080, USA. <sup>2</sup>Department of Neuroscience & McKnight Brain Institute, University of Florida, Gainesville, Florida 32611, USA. <sup>3</sup>Friday Harbor Laboratories, University of Washington, Friday Harbor, Washington 98250, USA. <sup>4</sup>Department of Biological Sciences, Auburn University, 101 Rouse Life Sciences, Auburn, Alabama 36849, USA. <sup>5</sup>Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain. <sup>6</sup>Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain. <sup>7</sup>Department of Psychiatry, Brudnick Neuropsychiatric Research Institute, University of Massachusetts Medical School, 303 Belmont Street, Worcester, Massachusetts 01604, USA. <sup>8</sup>Vavilov Institute of General Genetics, Russian Academy of Sciences (RAS), Gubkina 3, Moscow 119991, Russia. <sup>9</sup>Department of Chemistry and the Beckman Institute for Advanced Science and Technology, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, USA. <sup>10</sup>European Research Institute for the Biology of Ageing, University of Groningen Medical Center, Antonius Deusinglaan 1, Building 3226, Room 03.34, 9713 AV Groningen, The Netherlands. <sup>11</sup>Department of Medical Biophysics and Department of Immunology, University of Toronto, Sunnybrook Research Institute 2075 Bayview Avenue, Toronto, Ontario M4N 3M5, Canada. <sup>12</sup>Genetic Information Research Institute, 1925 Landings Dr., Mountain View, California 94043, USA. <sup>13</sup>Program in Molecular Medicine, University of Massachusetts Medical School, 222 Maple Avenue, Shrewsbury, Massachusetts 01545, USA. <sup>14</sup>Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK. <sup>15</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain. <sup>16</sup>Center for Brain Neurobiology and Neurogenetics and Institute of Cytology and Genetics, RAS, Lavrentyev Avenue, 10, Novosibirsk 630090, Russia. <sup>17</sup>Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Leninskiy Gory, 119991 Moscow, Russia.

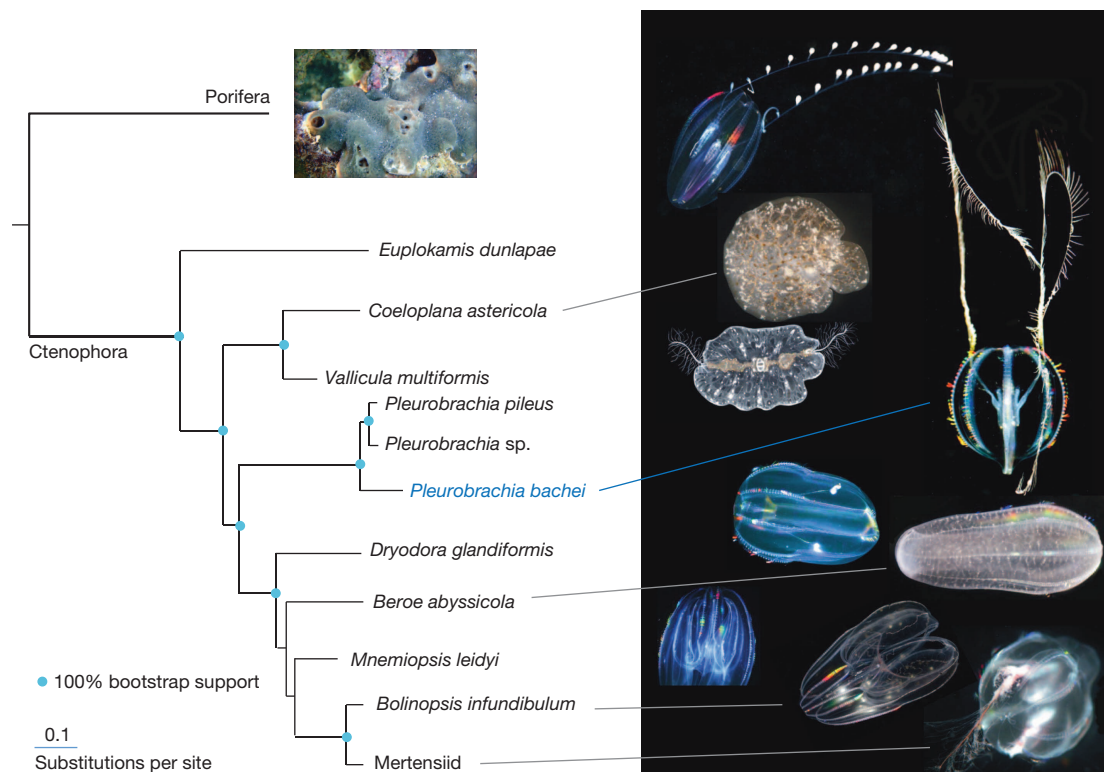




**Figure 1 | Ctenophores and their innovations.** **a**, The sea gooseberry, *Pleurobrachia bachei* (Supplementary Fig. 1), was selected as a target for genome sequencing due to preservation of traits ancestral for this lineage and because *in situ* hybridization/immunolabelling is possible. **b–e**, Major ctenophore innovations. **b**, Nervous system revealed by tyrosinated  $\alpha$ -tubulin immunolabelling (scale bar, 60  $\mu$ m). **c**, Scanning electron microscopy (SEM) imaging of nerve net in a tentacle pocket (scale bar, 20  $\mu$ m). **d**, Locomotory ciliated combs (SEM; scale bar, 100  $\mu$ m). **e**, Glue-secreting cells (colloblasts) in tentacles (SEM; scale bar, 20  $\mu$ m). **f**, Relationships among major animal clades with choanoflagellates sister to all Metazoa.

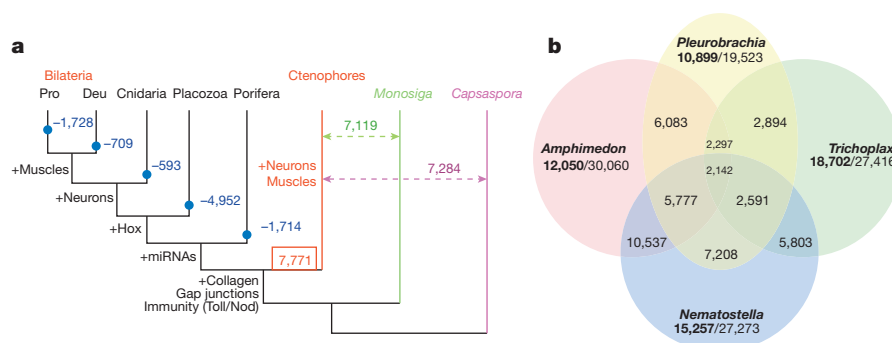
topology was unresolved. Weak support is probably due to underrepresentation of comparable transcriptomes from sponges and large protein divergence. Nevertheless, Shimodaira–Hasegawa tests based on expanded ctenophore sampling (with a reduced 114 gene matrix due to

lack of other ctenophore and sponge genomes; Supplementary Methods 7.2) also rejected Coelenterata but not Eumetazoa. Notably, relationships within Ctenophora were strongly supported (Fig. 2). Both cydippid and lobate ctenophores, previously viewed as monophyletic clades,



**Figure 2 | Phylogenomic reconstruction among major ctenophore lineages.** Cydippid (*Euplokamis*, *Pleurobrachia*, *Dryodora* and Mertensiidae) and lobate (*Mnemiopsis* and *Bolinopsis*) ctenophores were polyphyletic, suggesting

independent loss of both cydippid larval stage and tentacle apparatus as well as independent development of bilateral symmetry in benthic ctenophores, *Vallicula* and *Coeloplana* (Supplementary Data 4).



**Figure 3 | Gene gain and gene loss in ctenophores.** **a**, Predicted scope of gene loss (blue numbers; for example,  $-4,952$  in Placozoa) from the common metazoan ancestor. Red and green numbers indicate genes shared between bilaterians and ctenophores (7,771), as well as between ctenophores and other eukaryotic lineages sister to animals, respectively. Text on tree indicates

were recovered polyphyletic, indicating independent loss of both the cydippid larval stage and tentacle apparatus. Interestingly, *Platyctenida* was the second basal-most branch in the Ctenophore clade, suggesting that their benthic and bilateral nature are secondarily derived.

A highly reduced complement of animal-specific genes is a feature shared for the entire ctenophore lineage (Fig. 3 and Supplementary Table 15). HOX genes involved in antero-posterior patterning of body axes and present in all metazoans are absent in ctenophores and sponges<sup>18</sup> (Supplementary Tables 17 and 18). Likewise, canonical microRNA machinery (that is, *Drosha* and *Pasha*, Supplementary Table 19) is lacking in *Pleurobrachia* and other ctenophores. Using small RNA sequencing from *Pleurobrachia*, *Bolinopsis* and *Beroe*, we were unable to experimentally detect microRNAs (Supplementary Data 5.4). *Pleurobrachia* also lacks major elements that initiate innate immunity such as pattern recognition receptors (Toll-like, Nod-like, RIG-like, Ig-Toll-IL-1R domain) and immune mediators, MyD88 and RHD transcription factors, that are present in bilaterians, cnidarians and, in divergent forms, in sponges<sup>19,20</sup> (Supplementary Table 20).

Key bilaterian myogenic/mesoderm-specification genes are absent in *Pleurobrachia*'s genome and transcriptomes of ten other ctenophores (Supplementary Table 35). These data indicate that muscles<sup>21</sup> and, possibly, mesoderm evolved independently in Ctenophora to control the hydroskeleton, body shape and food capture. Thus, ctenophores might have independently developed complex phenotypes and tissue organization, raising questions about the nature of ctenophore-specific traits such as their unique development, combs, tentacles, aboral/apical organs and nervous systems.

## Ctenophore innovations

To assess the genomic bases of ctenophore-specific innovations, we performed RNA-seq profiling of the major developmental stages (Fig. 4a, b) as well as adult organs and identified genes differentially expressed in these structures. Many *Pleurobrachia* genes that have no homologues in other species are specifically expressed and most abundant during the 4- to 32-cell cleavage stages as well as in tentacles, combs and the aboral organ (Fig. 4b and Extended Data Fig. 4). Thus, structures that are known as ctenophore innovations (Fig. 1d, e) have the largest complement of highly expressed *Pleurobrachia*/ctenophore-specific genes. These data suggest extensive gene gain in cell lineages associated with early segregation of developmental potential leading to ctenophore-specific traits in structures controlling feeding, locomotion and integrative functions; a finding consistent with proposed 'orphan' genes contributing to variation in early development and evolution of novelties<sup>22,23</sup>.

Examples of known metazoan gene families that are considerably expanded in Ctenophora (Supplementary Data 5 and Supplementary Table 16) include collagens, RNA-editing enzymes and RNA-binding proteins (Supplementary Data 5). *Pleurobrachia*'s genome encodes the

emergence of complex animal traits and gene families. Deu, deuterostomes; Pro, protostomes. **b**, Uniquely shared and lineage-specific genes among basal metazoans. Values under species names indicate number of genes (bold) without any recognized homologues ( $e$ -value is  $10^{-4}$ ) versus the total number of predicted gene models in relevant species (Supplementary Table 14b).

highest number of RNA-editing enzymes (*ADAR1-4*, *ADAT1-3*, *CDA1-2*) reported so far among metazoans<sup>24,25</sup> (Supplementary Data 5.5), possibly acting as the generalized mechanism generating post-transcriptional diversity and ctenophore-specific traits in locomotory and integrative structures (combs and aboral organ). Matching the expansion of RNA regulatory mechanisms, *Pleurobrachia* has more RNA-binding proteins (especially RRM/ELAV, KH and NOVA<sup>26,27</sup>, Supplementary Table 21) than any basal metazoan or choanoflagellate examined. Dozens of RNA-binding proteins are selectively expressed and abundant during 8–64-cell stages (Supplementary Table 31), and might contribute to sequestration of RNAs and segregation in developmental potential leading to early cell-fate specification.

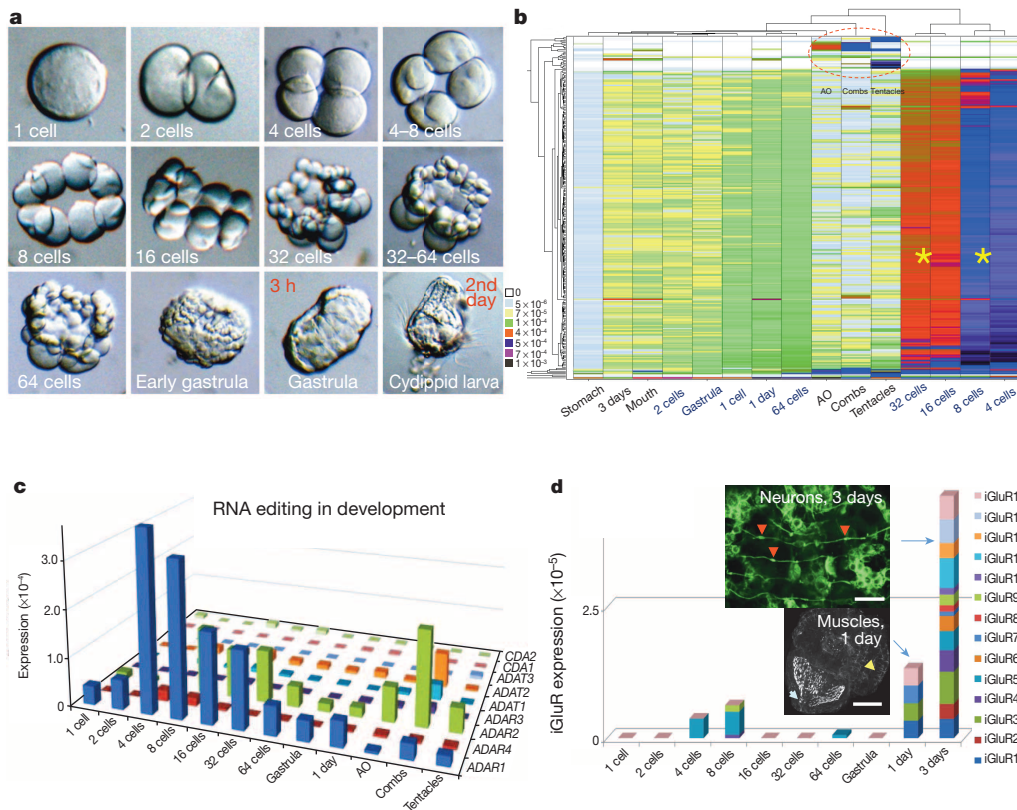
Phenotypic complexity positively correlates with presence and high differential expression of 92 homeodomain *Pleurobrachia* genes (Supplementary Data 5.2 and Supplementary Table 17)—76 genes have been reported in *Mnemiopsis*<sup>18</sup>, whereas the *Amphimedon* homeodomain complement consists of only 32 genes. In contrast, some developmental pathways are either absent (*Hedgehog*, *JAK/STAT*) or have reduced representation in ctenophores (TGF- $\beta$ , Wnt, Notch). Surprisingly, most Wnt genes are weakly expressed during *Pleurobrachia* development, whereas the ctenophore-specific subtype *WntX* is primarily restricted to adult neuroid elements such as polar fields, aboral organ and tentacular conductive tracts (Extended Data Fig. 5e), suggesting a distinct molecular make-up of neural systems.

## Parallel evolution of neural organization

Extensive parallel evolution of neural organization in ctenophores is the most evident. Compared to other animals with nervous systems, many genes controlling neuronal fate and patterning (for example, neurogenins, NeuroD, Achaete-scute, REST, HOX, Otx) are absent in the ctenophores that we sampled. Orthologues of pre- and postsynaptic genes also have limited representation (Supplementary Table 34), and they lack components (for example, neuroligin) critical for synaptic function in other eumetazoans.

Notably, our combined molecular, ultrasensitive metabolomic, immunohistochemical and pharmacological data strongly suggest that ctenophores do not use serotonin, acetylcholine, dopamine, noradrenaline, adrenaline, octopamine, histamine or glycine as intercellular messengers (Extended Data Figs 6 and 7g, Supplementary Data 5.8 and Supplementary Tables 22–26). Lack of ionotropic receptors for these molecules in ctenophores is consistent with this conclusion (Supplementary Table 26a). Most genes involved in the synthesis of neurotransmitters are absent in non-metazoan opisthokonts *Monosiga* and *Capsaspora*, suggesting that they are cnidarian/bilaterian innovations.

We next investigated the identity of the ctenophore transmitters. Physiological and pharmacological tests suggest that L-glutamate is a candidate neuromuscular transmitter in *Pleurobrachia* (Fig. 5b, Extended Data



**Figure 4 | Nature of ctenophore innovations.** **a**, Main developmental stages in *Pleurobrachia* from eggs to cleavage (2–64 cells), gastrulation (1–3 h) and formation of cydippid larvae (~24 h). **b**, Hierarchical clustering of approximately 400 ctenophore-specific genes differentially expressed among different developmental stages and adult structures as revealed by RNA-seq experiments. Colour index: black indicates highest level of expression, followed by purple, red then down to white indicating no expression. Most of these ctenophore-specific genes are primarily expressed during 4–32-cell stages (asterisks). The red circle indicates a subset of novel genes uniquely expressed in combs, tentacles and the aboral organ (AO). These genes lack recognized homologues in other organisms. **c**, Diversity and differential expression of RNA-editing genes in *Pleurobrachia* development and adult tissues (RNA-seq). *ADAR1* has the highest expression level in early cleavage stages whereas

*ADAR2* and *ADAR3* and *ADAT1* and *ADAT2* are most abundant in the combs. **d**, Morphological appearance of neurons during the third day of development (the top inset, neuronal cell bodies are stained with tyrosinated  $\alpha$ -tubulin antibodies, red arrows) correlates with abundant expression of multiple iGluR receptors, suggesting that glutamate has an important role as an intercellular messenger. Muscles formed well before neuronal differentiation at the end of the first day of development (the bottom inset, phalloidin staining, yellow arrow); white arrow points to the embryonic mouth with hundreds of cilia inside. In **c** and **d** expression levels of RNA-editing or iGluR genes are shown as a normalized frequency of sequence reads for a given transcript from all RNA-seq data for each developmental stage (Supplementary Methods). Scale bars in **d**: 15  $\mu$ m (top); 40  $\mu$ m (bottom).

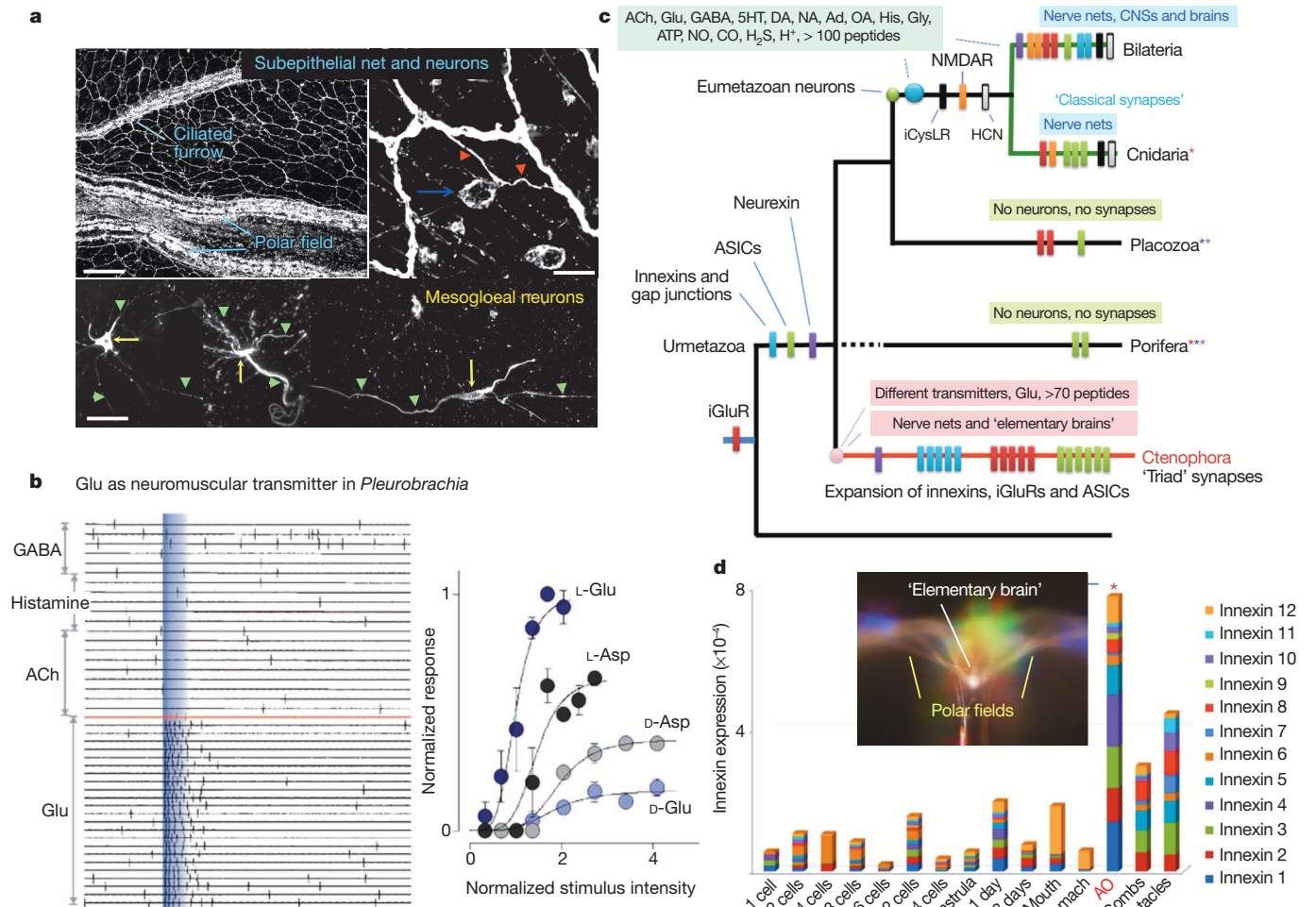
Fig. 7), able to induce rapid inward currents and raise intracellular  $\text{Ca}^{2+}$  concentration in muscle cells causing muscle contractions at nanomolar concentrations ( $10^{-7}$  M). In contrast, all other classical neurotransmitters were ineffective even in concentrations up to  $5 \times 10^{-3}$  M whereas D-glutamate as well as L-/D-aspartate have significantly less affinity in these assays (Fig. 5b).

The proposed role of glutamate as a signal molecule in ctenophores is supported by an unprecedented diversity of ionotropic glutamate receptors (iGluRs; Extended Data Fig. 7a, b and Supplementary Table 27)—far exceeding the number of genes encoding iGluRs in other basal metazoans<sup>28</sup>. iGluRs might have undergone a substantial adaptive radiation in Ctenophora as evidenced by unique exon/intron organization for many subtypes, and ctenophoran iGluRs form a distinct clade within the gene tree. Notably, during development, *Pleurobrachia*'s neurons are formed 2 days after the initial muscle formation, and first neurogenesis events correlate with co-expression of all iGlu receptors in hatching larvae (Fig. 4d). All cloned iGluRs also show remarkable cell-type-specific distribution with predominant expression in tentacles, followed by combs and the aboral organ, revealing well-developed glutamate signalling in adults (Extended Data Fig. 7b). Additionally, *Pleurobrachia* contains more genes for glutamate synthesis (8 glutaminases) and transport (8 sialins) than any other metazoan investigated<sup>29,30</sup>. Although we detected GABA ( $\gamma$ -aminobutyric acid, Supplementary Tables 22–24)

and immunohistochemically identified its localization in muscles, lack of pharmacological effects of GABA on *Pleurobrachia* behaviour and major motor systems, such as cilia, muscle and colloblasts, suggest that GABA is a by-product of glutamate metabolism by L-glutamic acid decarboxylase.

The first nervous systems are suggested to be primarily peptidergic in nature<sup>7</sup>. Although we did not find any previously identified neuropeptide homologue, secretory peptide prohormone processing genes (Supplementary Table 31) are present. We predicted 72 novel putative prohormones in *Pleurobrachia* and found >50 homologues in other sequenced ctenophores (Extended Data Fig. 8 and Supplementary Tables 28 and 32). Functions of these prohormone-derived peptides could include cell-to-cell signalling, toxins, involvement in innate immunity, or a combination of these functions. Several ctenophore-specific precursors are expressed in polarized cells around the mouth, tentacles and polar fields, indicative of a signalling role (Extended Data Fig. 8b). Secreted products of these prohormones may be natural ligands for >100 orphan neuropeptide-like G-protein-coupled receptors<sup>31</sup> identified in *Pleurobrachia* (Supplementary Table 26b). A second example of neuropeptide receptor candidates is amiloride-sensitive sodium channels (ASIC), which are also known to be regulated by different classes of short peptides and protons<sup>32</sup>. *Pleurobrachia*'s genome has 29 genes encoding ASICs—more than any organism sequenced so far—and expression of most correlated





**Figure 5 | Emergence of neural organization in *Pleurobrachia*.** **a**, Two neural nets in *Pleurobrachia* as revealed by tyrosinated  $\alpha$ -tubulin immunostaining. Top image shows subepithelial net with concentrations of neuronal elements in the polar fields and ciliated furrows, known as structures involved in sensory and motor functions, respectively (blue arrow in right insert indicates location of a neuronal somata with individual neurites marked by red arrows). The bottom image shows neurons of the mesogloal net (arrows are neuronal somata; arrowheads are neuronal processes). Note that phalloidin (a muscle marker) did not stain these cells. Scale bar, 120  $\mu$ m (top); 10  $\mu$ m (bottom images). **b**, L-glutamate ( $10^{-7}$ – $10^{-3}$  M) induced action potentials in muscle cells, whereas other transmitter candidates were ineffective even at concentrations up to 5 mM. Typical responses of ctenophore muscle cells to local pulses of a transmitter application were recorded both as individual action potentials (whole-cell current-clamp mode) and video contractions from a single muscle cell. The graph shows normalized responses from the same muscle cell, indicating that L-glutamate is the most potential excitatory

molecule compared to D-glutamate or L-/D-aspartate (Supplementary Methods). ACh, acetylcholine. **c**, Key molecular innovations underlying neural organization in ctenophores. Bars indicate the presence or relative expansions of selected gene families in all basal metazoan lineages from the inferred urmetazoan ancestor. The data suggest that sponges and placozoans never developed neural systems, or, unlikely assuming the presence of neuronal organization in the urmetazoan ancestor, sponges and placozoans lost their nervous systems. Either hypothesis points towards extensive parallel evolution of neural systems in ctenophores versus the Bilateria + Cnidaria clade. **d**, The aboral organ has the greatest diversity and highest expression levels of 12 gap junction proteins, suggesting unmatched expansion of electrical signalling in this complex integrative organ—an analogue of an elementary brain in ctenophores. Expression of different innexins is shown as a summation of normalized frequencies of respective sequencing reads in RNA-seq data obtained from each developmental stage and adult tissues (y axis is 'expression frequency' or 'expression'; see Supplementary Methods).

with developmental appearance of neurons (Supplementary Table 31). ASIC expression is most abundant in tentacles, combs and aboral organs—structures enriched in neural elements and under complex synaptic control.

Moreover, ctenophores evolved an enormous diversity of electrical synapses (absent in *Nematostella*, *Amphimedon* and *Trichoplax*) with 12 gap junction proteins (pannexins/innexins<sup>33</sup> but not chordate-specific connexins) found in *Pleurobrachia*. All pannexins/innexins have their highest expression in the aboral organ followed by tentacles and combs (Fig. 5d). Non-metazoans lack pannexin orthologues, suggesting that these are metazoan innovations with profound expansion of this family in ctenophores. The aboral organ, combs and tentacles have a relatively large diversity of ion channels (Extended Data Fig. 9b), confirming complex regulation of excitability in these structures. However, the overall

complement of voltage-gated ion channels in ctenophores is reduced compared to other eumetazoans<sup>34</sup> (Extended Data Fig. 9a).

Our genome-wide survey also indicates that some bilaterian and cnidarian pan-neural markers are present (for example, three *elav* and *musashi* genes), but they are not expressed in neurons; a finding consistent with early divergence and extreme parallel evolution of neural systems in this lineage (Extended Data Figs 5 and 9b).

## Discussion

Figure 5c summarizes key molecular innovations underlying neural organization in ctenophores. Evidently, with a markedly different molecular and genomic make-up, ctenophores have achieved complex phenotypic plasticity and tissue organization. Thus, ctenophores might represent

remarkable examples of convergent evolution including the emergence of neuro-muscular organization from the metazoan common ancestor without differentiated nervous system or bona fide neurons (Extended Data Fig. 10b and Supplementary Data 15). The alternative 'single-origin hypothesis', where the common ancestor of all metazoans had a nervous system with complex molecular and transmitter organization including all classical cnidarian/bilaterian transmitters and neurogenic genes (Extended Data Fig. 10a), is a less parsimonious scenario. This hypothesis implies that ctenophores, despite being active predators, underwent massive loss of neuronal and signalling toolkits and then replaced them with novel neurogenic and signalling molecules and receptors.

These findings are of relevance for regenerative and synthetic biology in designing novel signalling pathways and systems. In this case, ctenophores and their genomes present matchless examples of 'experiments' in nature and the possible preservation of ancient molecular toolkits lost in other animal lineages.

## METHODS SUMMARY

Genome sequencing and transcriptome profiling used high-throughput methods. Gene prediction (Augustus/Egenesh) and annotations were complemented by metabolome and microchemical assays using capillary electrophoresis with laser-induced fluorescence and mass spectrometry detectors to validate the presence/absence of selected transmitters. RNA-seq, PCR with reverse transcription (RT-PCR), *in situ* hybridization and immunohistochemistry were used to characterize and localize gene expression in adults and embryos. Pharmacological and electrophysiological tests were performed on semi-intact preparations using freshly isolated ctenophore cells.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 September 2013; accepted 23 April 2014.

Published online 21 May; corrected online 4 June 2014 (see full-text HTML version for details).

- Nielsen, C. *Animal Evolution: Interrelationships of the Living Phyla* 402 (Oxford Univ. Press, 2012).
- Hernandez-Nicaise, M.-L. in *Microscopic Anatomy of Invertebrates: Placozoa, Porifera, Cnidaria, and Ctenophora* Vol. 2 (eds Harrison, F. W. & Westfall, J. A.) 359–418 (Wiley, 1991).
- Tamm, S. L. in *Electrical Conduction and Behavior in "Simple" Invertebrates* 266–358 (Clarendon, 1982).
- Tang, F., Bengtson, S., Wang, Y., Wang, X. L. & Yin, C. Y. *Eoandromeda* and the origin of Ctenophora. *Evol. Dev.* **13**, 408–414 (2011).
- Dzik, J. Possible ctenophoran affinities of the Precambrian "sea-pen" Rangea. *J. Morphol.* **252**, 315–334 (2002).
- Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature Rev. Mol. Cell Biol.* **14**, 341–356 (2013).
- Moroz, L. L. On the independent origins of complex brains and neurons. *Brain Behav. Evol.* **74**, 177–190 (2009).
- Moroz, L. L. Phylogenomics meets neuroscience: how many times might complex brains have evolved? *Acta Biol. Hung.* **63** (suppl. 2), 3–19 (2012).
- Pennisi, E. Nervous system may have evolved twice. *Science* **339**, 391 (2013).
- Nosenko, T. et al. Deep metazoan phylogeny: When different genes tell different stories. *Mol. Phylogenet. Evol.* **67**, 223–233 (2013).
- Dunn, C. W. et al. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745–749 (2008).
- Hejnol, A. et al. Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc. R. Soc. Lond. B* **276**, 4261–4270 (2009).
- Philippe, H. et al. Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19**, 706–712 (2009).
- Pick, K. S. et al. Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. *Mol. Biol. Evol.* **27**, 1983–1987 (2010).
- Schierwater, B. et al. Concatenated analysis sheds light on early metazoan evolution and fuels a modern "Urmetazoon" hypothesis. *PLoS Biol.* **7**, e1000020 (2009).
- Kohn, A. B. et al. Rapid evolution of the compact and unusual mitochondrial genome in the ctenophore, *Pleurobrachia bachei*. *Mol. Phylogenet. Evol.* **63**, 203–207 (2012).
- Shimodaira, H. & Hasegawa, M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**, 1114–1116 (1999).
- Ryan, J. F., Pang, K., Mullikin, J. C., Martindale, M. Q. & Baxevanis, A. D. The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. *Evodevo* **1**, 9 (2010).
- Lange, C. et al. Defining the origins of the NOD-like receptor system at the base of animal evolution. *Mol. Biol. Evol.* **28**, 1687–1702 (2011).
- Leulier, F. & Lemaitre, B. Toll-like receptors—taking an evolutionary approach. *Nature Rev. Genet.* **9**, 165–178 (2008).

- Steinmetz, P. R. et al. Independent evolution of striated muscles in cnidarians and bilaterians. *Nature* **487**, 231–234 (2012).
- Chen, S., Krinsky, B. H. & Long, M. New genes as drivers of phenotypic evolution. *Nature Rev. Genet.* **14**, 645–660 (2013).
- Tautz, D. & Domazet-Loso, T. The evolutionary origin of orphan genes. *Nature Rev. Genet.* **12**, 692–702 (2011).
- Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* **79**, 321–349 (2010).
- Savva, Y. A., Rieder, L. E. & Reenan, R. A. The ADAR protein family. *Genome Biol.* **13**, 252 (2012).
- Yano, M., Hayakawa-Yano, Y., Mele, A. & Darnell, R. B. Nova2 regulates neuronal migration through an RNA switch in disabled-1 signaling. *Neuron* **66**, 848–858 (2010).
- Zhang, C. et al. Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* **329**, 439–443 (2010).
- Traynelis, S. F. et al. Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev.* **62**, 405–496 (2010).
- Omote, H., Miyaji, T., Juge, N. & Moriyama, Y. Vesicular neurotransmitter transporter: bioenergetics and regulation of glutamate transport. *Biochemistry* **50**, 5558–5565 (2011).
- El Mestikawy, S., Wallen-Mackenzie, A., Fortin, G. M., Descarries, L. & Trudeau, L. E. From glutamate co-release to vesicular synergy: vesicular glutamate transporters. *Nature Rev. Neurosci.* **12**, 204–216 (2011).
- Palczewski, K. & Orban, T. From atomic structures to neuronal functions of G protein-coupled receptors. *Annu. Rev. Neurosci.* **36**, 139–164 (2013).
- Krishtal, O. The ASICs: signaling molecules? Modulators? *Trends Neurosci.* **26**, 477–483 (2003).
- Panchin, Y. V. Evolution of gap junction proteins—the pannexin alternative. *J. Exp. Biol.* **208**, 1415–1419 (2005).
- Liebkekind, B. J., Hillis, D. M. & Zakon, H. H. Evolution of sodium channels predates the origin of nervous systems in animals. *Proc. Natl Acad. Sci. USA* **108**, 9154–9159 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank Friday Harbor Laboratories for facilities during animal collection and Marine Genomics apprenticeships (L.L.M., B.J.S.); E. Dabe, G. Winters, J. Netherton, N. Churches and C. Bostwick for help with animal, tissue, *in situ*, RNA and DNA assays; and X.-X. Tan, F. Lu and T. Tyazelova for sequencing. We thank F. Nivens for videos and P. L. Williams for database support. This work was supported by NSF (NSF-0744649 and NSF CNS-0821622 to L.L.M.; NSF CHE-1111705 to J.V.S.), NIH (1R01GM097502, R01MH097062, R21RR025699 and 5R21DA030118 to L.L.M.; P30 DA018310 to J.V.S.; R01 AG029360 and 1S1ORR027052 to E.I.R.), NASA NNX13AJ31G (to K.M.H., L.L.M. and K.M.K.), NSERC 458115 and 211598 (J.P.R.), University of Florida Opportunity Funds/McKnight Brain Research and Florida Biodiversity Institute (L.L.M.), Rostock Inc./A.V. Chikunov (E.I.R.), grant from Russian Federation Government 14.B25.31.0033 (Resolution No.220) (E.I.R.). F.A.K., I.S.P. and R.D. were supported by HHMI (55007426), EMBO and MINECO (BFU2012-31329 and Sev-2012-0208). Contributions of AU Marine Biology Program 117 and Molette laboratory 22.

**Author Contributions** L.L.M. conceived the project, designed the experiments and wrote the manuscript. A.B.K., A.P.G., D.R., E.K., F.G., R.S., T.P.M., E.I.R. and L.L.M. prepared gDNA, RNA samples and performed sequencing; I.S.P., F.A.K., V.V.S., F.Y., M.R.C., A.B.K. and L.L.M. did assemblies, gene model prediction and annotations; K.M.K. and K.M.H. performed phylogenomic analysis; A.P., A.B.K. and L.L.M. worked on gene family gain/loss analysis; F.A.K. and R.D. characterized protein divergence; A.B.K., L.L.M. and K.M. characterized homeodomain complement; S.D., C.D., J.V.S. and L.L.M. performed capillary electrophoresis/microchemical metabolomic assays; A.P.G., A.B.K., E.B. and E.I.R. did small RNA sequencing and analysis; K.M.B. and J.P.R. characterized immune gene complement; V.V.K. and J.J. characterized transposons; T.P.N. and L.L.M. performed immunolabelling, electron microscopy and pharmacological assays; Y.V.B. and L.L.M. performed pharmacological, electrophysiological and imaging assays on muscles; D.O.G., M.R.C., A.B.K. and L.L.M. performed secretory peptide prediction; A.B.K. and L.L.M. analysed RNA-seq data; A.B.K. performed methylation analysis; B.J.S., A.B.K. and L.L.M. analysed developmental data; J.J.S., D.O.G., Y.B., R.B., A.F., A.B.K. and L.L.M. performed *in situ* hybridization experiments; C.E.M. identified species and wrote their description and biology; all authors contributed to preparation of the manuscript and the text.

**Author Information** The genome and transcriptome data are available at DDBJ/EMBL/GenBank AVPN000000000; the BioProject PRJNA213480 and SRA SRP001155; and (<http://neurobase.rc.ufl.edu/Pleurobrachia>). Principal Investigator: Leonid L. Moroz. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.L.M. ([moroz@whitney.ufl.edu](mailto:moroz@whitney.ufl.edu)), E.I.R. ([Evgeny.Rogaev@umassmed.edu](mailto:Evgeny.Rogaev@umassmed.edu), gDNA-seq) or A.B.K. ([abkohn@msn.com](mailto:abkohn@msn.com), RNA-seq).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>



## METHODS

**Source material.** Animals (*Pleurobrachia bachei*, *Euplokamis dunlapae*, *Dryodora glandiformis*, *Beroe abyssicola*, *Bolinopsis infundibulum* and *Mertensii*) were collected at Friday Harbour Laboratories (Pacific North-Western Coast of USA) and maintained in running seawater for up to 2 weeks. Other species were collected at the Atlantic coast of Florida and around Woods Hole, Massachusetts (*Pleurobrachia pileus*, *Pleurobrachia* sp., *Mnemiopsis leidyi*) as well as central Pacific (Palau, Hawaii, *Coeloplana astericola*, *Vallicula multiformis*). Animals were anaesthetized in 60% (volume/body weight) isotonic  $MgCl_2$  (337 mM). Specific tissues were surgically removed with sterile fine forceps and scissors and processed for DNA/RNA isolations as well as metabolomics or pharmacological/electrophysiological tests. Whole animals were used for all *in situ* hybridization and immunohistochemical tests as described<sup>35</sup>. Genomic DNA (gDNA) was isolated using Genomic-tip (QIAGEN) and total RNA was extracted using RNAqueous-Micro (Ambion/Life Technology) or RNAqueous according to manufacturers' recommendations. Quality and quantity of gDNA was analysed on a Qubit2.0 Fluorometer (Life Technologies) and for RNA we used a 2100 Bioanalyzer (Agilent Technologies). For all details see Supplementary Methods sections 1.1–1.3.

**Genome sequencing.** All genomic sequence data for *de novo* assembly were generated on Roche 454 Titanium and Illumina Genome Analyzer IIx, HiSeq2000 and MiSeq instruments using both shotgun pair-end and mate-pair sequencing libraries with 3–9 kb inserts as summarized in Supplementary Tables 1 and 2. Shotgun sequencing was performed from a single individual. Owing to a limited amount of starting gDNA, mate pair libraries were constructed from 10–12 individuals. In total, the genome sequencing is composed of 132,015,600,107 bp or ~132 Gb of data, which corresponds to 733–825× physical coverage of the *Pleurobrachia* genome (the size of the *P. bachei* genome is estimated to be ~160–180 Mb); see Supplementary Methods sections 1.4–2.1.2.

**Genome assemblies.** The *Pleurobrachia bachei* draft genome was assembled using a custom approach designed to leverage the individual strengths of three popular *de novo* assembly packages and strategies: Velvet<sup>36</sup>, SOAPdenovo<sup>37</sup>, and pseudo-454 hybrid assembly with ABySS<sup>38</sup>. First, using filtered and corrected data, we performed individual assemblies from 454 and Illumina reads by the Newbler (Roche, Inc.) software. Then the merged/hybrid assembly was achieved using three individual assemblies (SOAPdenovo, Velvet and ABySS/Newbler as described in Supplementary Methods 2.2). Three gene model predictions were performed by Augustus<sup>39</sup> and Fgenesh predictions with the Softberry Inc. Fgenesh++ pipeline<sup>40,41</sup> to incorporate information from full-length cDNA alignments and similar proteins from the eukaryotic section of the NCBI NR database<sup>42</sup>. After initial gene predictions in each of the three sets of genomic scaffolds, we screened each set of gene models for internal redundancy with the BLASTP program from NCBI's BLAST+ software suite<sup>43</sup>. A model was considered redundant if it: had 90% identity to other model; the alignment between the two models had a bit score of at least 100; and the model was shorter than the other model.

Scaffolds producing these gene models were pooled and then screened for prokaryotic contamination using UCSC's BLAT software package<sup>44</sup> to produce the draft genome assembly version 1.0 (statistics can be found in Supplementary Table 5 and Supplementary Methods 2).

**Genome annotation.** For annotation, gene models were uploaded to the In-VIGO BLAST interface, a blastp alignment of gene models was performed against the entirety of NCBI's non-redundant protein database and the Swiss-Prot protein database, and subsequently annotated in terms of Gene Ontology and KEGG pathways as well as Pfam domain identification. Transposable elements (TEs) were identified using not only WU-BLAST and its implementation in CENSOR but also databases for all known classes, superfamilies and clades of TEs described in the literature and/or collected in Repbase<sup>45</sup>. Detected sequences have been clustered based on their pairwise identities by using BLASTclust. All autonomous non-LTR retrotransposons have been classified based on RTclass1 (ref. 46). To merge partially predicted, non-redundant gene models with assembled transcriptome data, a custom Java tool was developed. This Java tool extended partial gene model predictions based on using transcriptome sequences to bridge 5' and 3' fragments of partially predicted genes. Using this Java tool, analysis of alignments of non-redundant gene models to assembled *Pleurobrachia* transcriptomes resulted to 19,523 (Supplementary Table 30) gene models. These gene models were used to also identify their possible homologues in assembled transcriptomes from 10 other ctenophore species sequenced (Supplementary Tables 10 and 11). All genomic sequences were submitted to NCBI on SRA accession number Project SRP001155 (Supplementary Methods 3.1–3.2).

**Transcriptome sequencing and annotation.** Three sequencing technology platforms were used for transcriptome profiling (RNA-seq): Roche 454 Titanium, Illumina HiSeq2000 and Ion Proton/PGM (Ion Torrent, Life Technologies). RNA-seq was performed from all major embryonic and developmental stages (1 cell, 2 cells, 4 cells, 8 cells, 16 cells, 32 cells, 64 cells, early and later gastrula, 1 day and 3 day larvae), major adult tissues and organs (combs, mouth, tentacles, stomach, the aboral

organ, body walls), and whole body of *Pleurobrachia bachei*. We developed a reduced representation sequencing protocol for the 454 and Ion Torrent sequencing platforms that can detect low abundance transcripts<sup>47</sup>. The method reduces the amount of sequencing and gives more accurate quantification and additional details of the procedure are reported elsewhere<sup>47,48</sup>. In summary, we have generated 499,699,347 reads or ~47.9 Gb to achieve approximately 2,000× coverage of the *Pleurobrachia* transcriptome.

In addition, Illumina HiSeq sequencing was also performed with RNA extracted from the following ctenophore species: *Euplokamis dunlapae*, *Coeloplana astericola*, *Vallicula multiformis*, *Pleurobrachia pileus*, *Pleurobrachia* sp. (collected from the Middle Atlantic and later identified as a subspecies of *P. pileus*), *Dryodora glandiformis*, *Beroe abyssicola*, *Mnemiopsis leidyi*, *Bolinopsis infundibulum* and an undescribed species which belongs to the family Mertensidae (Supplementary Table 3). Each sequencing project was individually assembled using the Trinity *de novo* assembly package<sup>49</sup> and in selected cases using MIRA. Reads from developmental stages were also assembled using the CLCBio Genomics Workbench. Before each assembly, reads were quality trimmed and had adaptor contamination removed with cutadapt<sup>50</sup>. Full summaries of the transcriptome assemblies are presented in Supplementary Tables 4 and 10. Each transcriptome was mapped to the *Pleurobrachia* genome, and aligned to both NCBI's non-redundant protein database (NR) and the UniProtKB/Swiss-Prot (SP) protein database. Gene Ontology<sup>51</sup> and Kyoto Encyclopedia of Genes and Genomes<sup>52,53</sup> (KEGG) terms were associated with each transcript. By first translating transcripts in all six reading frames, Pfam/SMART domains<sup>54</sup> were assigned to each reference transcriptome.

Each reference transcriptome and its full set of annotation and expression data was uploaded to our transcriptome database <http://moroz.hpc.ufl.edu/slimebase2/browse.php> for downstream analysis and visualization<sup>55,56</sup>. The database is integrated with UCSC type genome browser. Via the genome project homepage (<http://neurobase.rc.ufl.edu/Pleurobrachia>) all data sets have direct download options. Quantification of gene expression profiling was performed on all transcriptional data as described in Supplementary Methods 4.4. Hierarchical clustering was performed by Spotfire agglomerative algorithm. All primary transcriptome data was submitted to NCBI on SRA accession number Project SRP000992. (See Supplementary Methods 4.1–4.2.3 for details.)

**Phylogenetic analyses.** To reconstruct basal metazoan phylogeny (see controversies in<sup>10–15,57</sup>), we conducted two sets of phylogenomic analysis using tools described elsewhere<sup>58</sup>. All analyses included new data from *Pleurobrachia bachei* and the sponges *Sycon* (Calcarea) and *Aphrocallistes* (Hexactinellida). For the first set of analyses, Ctenophora was represented by two species of *Pleurobrachia* and *Mnemiopsis leidyi*. Initial analyses included the taxa in Supplementary Table 12. For a subsequent analysis, sampling within Ctenophora was expanded to include ten additional taxa, each represented by a relatively deeply sequenced Illumina transcriptome (Supplementary Table 13). In order to reduce noise in the phylogenetic signal, we used strict criteria to exclude paralogues, highly derived sequences, mistranslated sequence regions, and ambiguously aligned positions in sequence alignments. Analyses were conducted in RAxML 7.2.7 (refs 59) using maximum likelihood (ML) with the CAT + WAG + F model. Topological robustness (that is, nodal support) for all ML analyses was assessed with 100 replicates of nonparametric bootstrapping. Details of phylogenomic analyses are presented in Supplementary Methods 7. Shimodaira–Hasegawa test<sup>17</sup> was implemented in RAxML with the PROTGAMMAWAG model<sup>17</sup>.

In order to examine evolution of single genes or gene families, alignments were performed with either ClustalX2<sup>60–62</sup> or Muscle<sup>63</sup> then, if appropriate, either trimmed manually or trimmed using GBLOCKS<sup>64</sup> to exclude ambiguously aligned positions. Once alignments were obtained, gene trees were reconstructed in MEGA 5<sup>65</sup> using ML with the Whelan and Goldman (WAG) model. The bootstrap consensus tree was inferred from 100 replicates. All positions containing gaps and missing data were eliminated. Pfam composition<sup>54</sup>, Gene Ontology<sup>51</sup> and KEGG<sup>52,53</sup> were used to further validate *P. bachei* orthologues. Analyses of gene gain and gene loss were performed using custom scripts as described elsewhere<sup>66</sup> and in Supplementary Methods 7.

**Analysis of DNA methylation.** ELISA-based colorimetric assays (Epigentech) were performed to quantify both global 5-mC and 5-hmC methylation in the *P. bachei* genome. A total of 6 individual *P. bachei* and three rat (positive control) were used (Supplementary Methods 1.2). Three biological and technical replicates were performed for every sample. Absolute quantification of 5-mC and 2-hmC were determined and data is reported as a mean ± s.e.m. (Supplementary Methods 8).

**Molecular cloning, *in situ* hybridization and immunohistochemistry.** Methods were similar as reported elsewhere<sup>35,47,48,67</sup> with some modifications (Supplementary Methods 9–11).

**Scanning electron microscopy.** Animals were fixed in 2.5% glutaraldehyde in 0.2 M phosphate-buffered saline (pH 7.6) for 3–4 h at room temperature, and washed. For secondary fixation, we used 2% osmium tetroxide in 1.25% sodium bicarbonate for 2–3 h at room temperature. After dehydration in ethanol, samples were placed

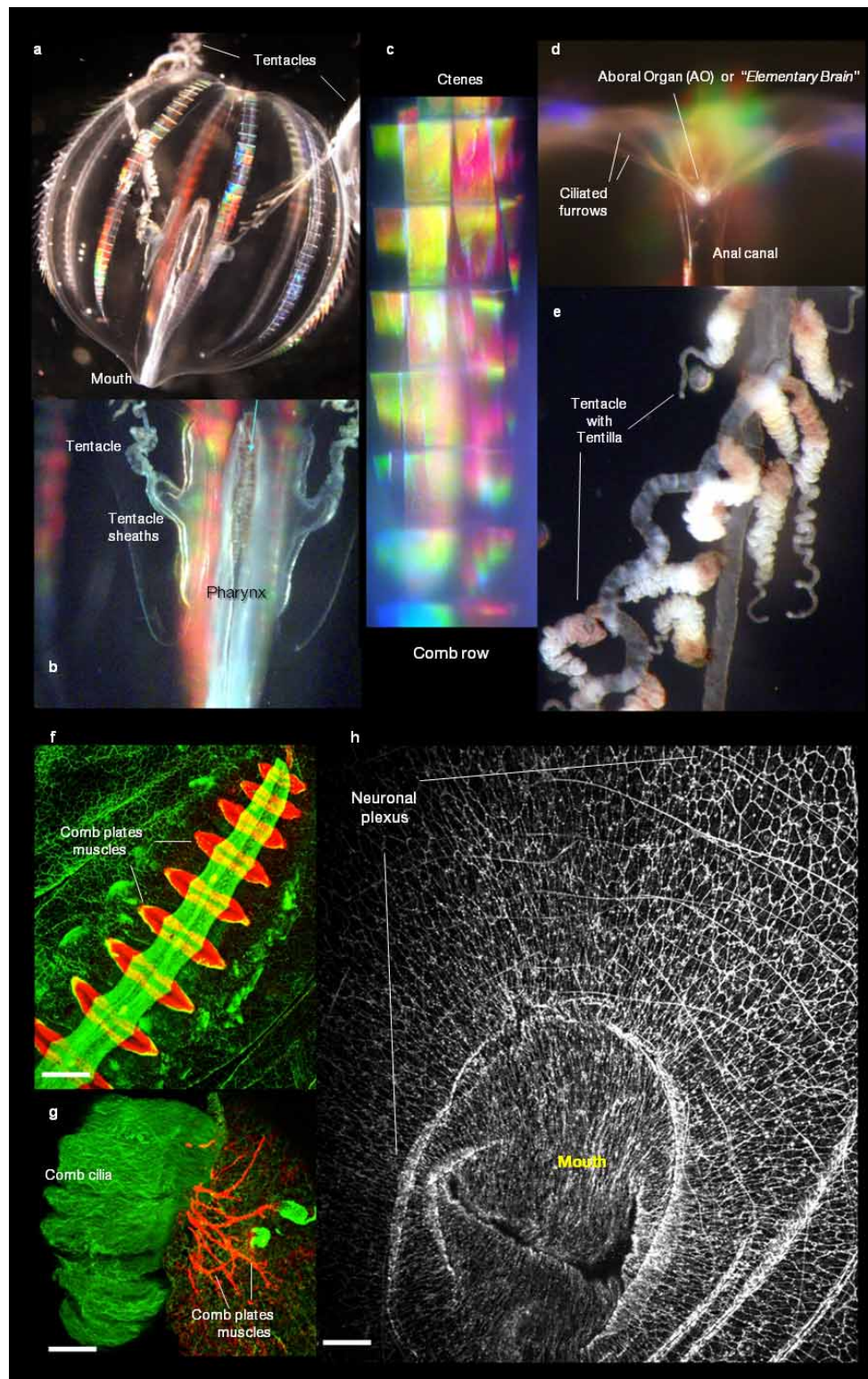


for drying in Samdri-790 Critical Point Drying. After drying the samples were coated on Sputter Coater. SEM observations and recordings were done on NeuScope JCM-5000 microscope (Supplementary Methods 12).

**Electrophysiological methods, calcium imaging and pharmacological assays.** Patch electrodes for extracellular and whole-cell recordings were pulled from borosilicate capillary (P-87, Sutter Instruments). All currents were recorded using an Axopatch or 200B amplifier controlled by a Digidata 1322A and pClamp 9.2. Action potentials (APs, spikes) were recorded in track mode using cell-attached loose-patch configuration. Whole-cell currents were recorded in voltage-clamp mode at a holding potential of  $-70$  mV. Neurotransmitter candidate (see Supplementary Method 15) application for both extracellular AP and whole cell recordings were performed with a rapid solution changer, RSC-160 (Bio-Logic-Science Instruments, France). Data were analysed with Clampfit 9.0 (Molecular Devices) in combination with SigmaPlot 10.0. Videomicroscopy and time-lapse series were acquired with QImaging EXi CCD camera using DIC mode of Nikon Eclipse 2000 inverted microscope. Calcium imaging was performed on isolated ctenophore muscle cells using an Olympus IX-71 inverted microscope equipped with a cooled CCD camera (ORCA R2, Hamamatsu). Cells were injected with calcium-sensitive probe (Fluo-4,  $\sim 5$   $\mu$ M) through a patch pipette. Fluorescence imaging was performed under the control of Imaging Workbench 6 software. Stored time series image stacks were analysed off-line using Imaging Workbench 6, Clampfit 10.3, SigmaPlot 10/11 or exported as TIFF files into ImageJ 1.42. Pharmacological tests and behavioural assays with video recording were performed on intact animals in 5–40 l aquaria or on semi-intact preparations in a Sylgard-coated Petri dish with free cilia beating and muscle contractions. To monitor and quantify cilia movements we used glass microelectrodes filled with 2 M potassium acetate with resistances of 5–20 M $\Omega$  with electrical signals recorded by A-M System amplifiers (Neuroprobe 1600) and Gould Recorder (WindoGraf 980). **Determination of the presence of classical neurotransmitters by capillary electrophoresis (CE).** Two CE separation techniques were used to analyse tissue extracts for the presence of a number of neurotransmitters (Supplementary Tables 22 and 23; Supplementary Methods 17). While both methods used CE separations, complimentary detection methods, laser-induced native fluorescence (LINF)<sup>68</sup> and electrospray ionization mass spectrometry (ESI-MS)<sup>69,70</sup> were used to ensure broad coverage and low detection limits for the specific analytes of interest. Whole bodies of small animals as well as individual organs and tissues were removed, rinsed with ultrapure water and analytes were extracted using 49.5/49.5/1, methanol (LC-MS grade)/water/glacial acetic acid (99%) by volume, homogenized, centrifuged and supernatant was removed and frozen at  $-80^{\circ}\text{C}$  until analysis. The CE-LINF instrument used ultraviolet excitation at 264 nm and the native fluorescence emission was collected and recorded using a UV-enhanced CCD array (Spec-10; 2KBV/LN; Princeton Instruments). CE separations were performed by hydrodynamic injection of 10 nl of sample and using 25 mM citric acid (pH 2.5, applied voltage +30 kV) or 50 mM borate (pH 9.5, applied voltage +21 kV). Analytes were identified based on comparison of both the migration time and fluorescence spectrum to that of standard mixtures of analytes. CE-ESI-MS analysis was performed using a Bruker Microtof or a Maxis (Bruker Daltonics) mass spectrometer for detection. All separations were performed using 1% formic acid in water as the electrolyte and applied voltage of +30 kV. Sheath liquid was 0.1% formic acid in 50/50 methanol/water. Samples were hydrodynamically injected for a total volume of  $\sim 6$  nl. Mass spectra were collected and recorded at a rate of 2 Hz with calibration performed using sodium formate clusters. Analytes were identified based on comparison of both the CE migration time and mass match to that of standard mixtures of analytes.

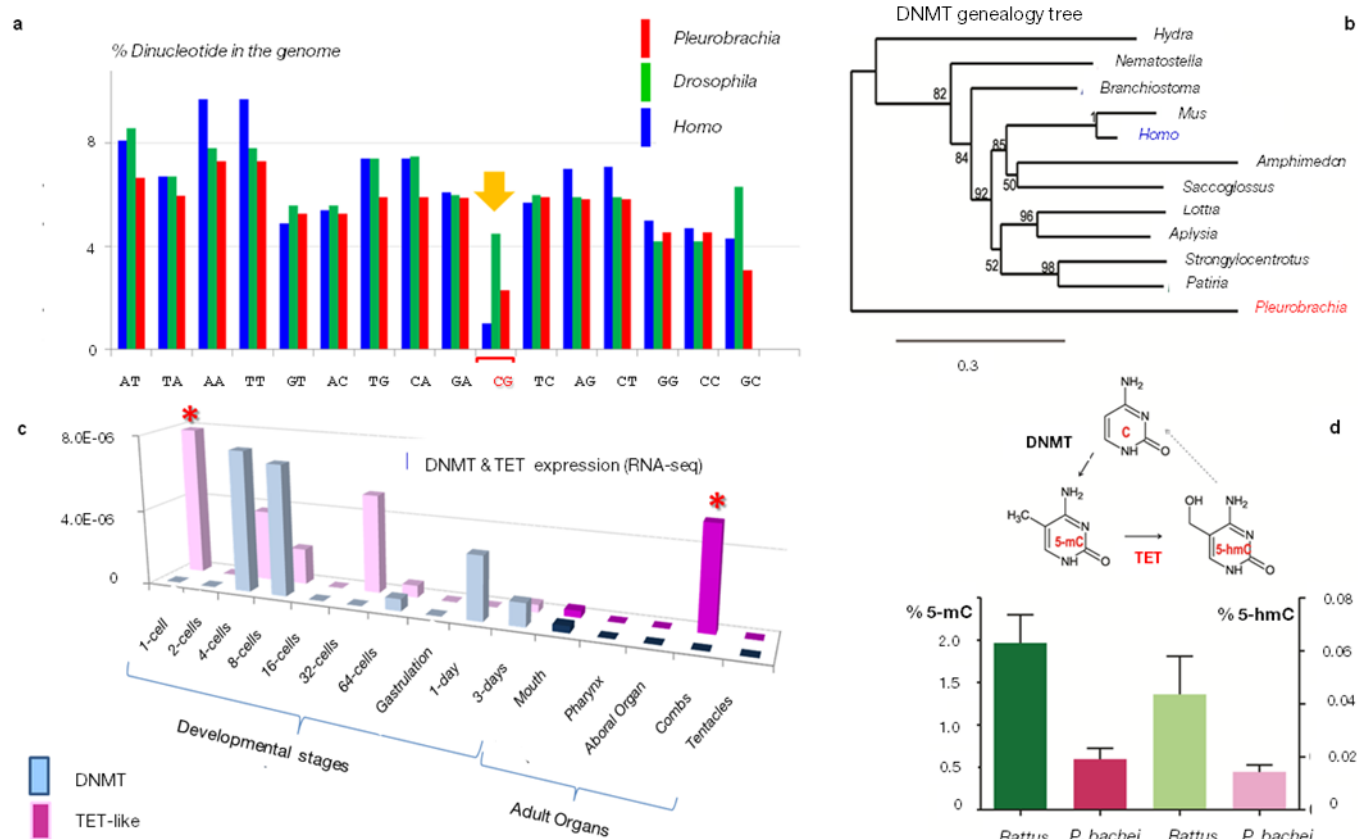
35. Moroz, L. L. & Kohn, A. B. Analysis of gene expression in neurons and synapses by multi-color *in situ* hybridization. *Methods Mol. Biol.* (in the press).
36. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
37. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 1–6 (2012).
38. Simpson, J. T. *et al.* ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123 (2009).
39. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).

40. Salamov, A. A. & Solovyev, V. V. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516–522 (2000).
41. Solovyev, V. in *Handbook of Statistical Genetics* (eds Balding, D. J., Bishop, M. & Cannings, C.) 97–159 (John Wiley & Sons, 2007).
42. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **37**, D5–D15 (2009).
43. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–4002 (1997).
44. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
45. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
46. Kapitonov, V. V., Tempel, S. & Jurka, J. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. *Gene* **448**, 207–213 (2009).
47. Kohn, A. B., Moroz, T. P., Barnes, J. P., Netherton, M. & Moroz, L. L. Single-cell semiconductor sequencing. *Methods Mol. Biol.* **1048**, 247–284 (2013).
48. Moroz, L. L. & Kohn, A. B. Single-neuron transcriptome and methylome sequencing for epigenomic analysis of aging. *Methods Mol. Biol.* **1048**, 323–352 (2013).
49. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol.* **29**, 644–652 (2011).
50. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet* **17**, 10–12 (2011).
51. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
52. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
53. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–D114 (2012).
54. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141 (2004).
55. Girardo, D. O., Citarella, M. R., Kohn, A. B. & Moroz, L. L. Automatic transcriptome analysis and quest for signaling molecules in basal metazoans. Int. Comp. Biol. Abstract meeting P1.136 (2012).
56. Girardo, D. O., Citarella, M., Kohn, A. B. & Moroz, L. L. Zero-click, automatic assembly, annotation and visualization workflow for comparative analysis of transcriptomes: The quest for novel signaling pathways. Int. Comp. Biol. 11.6 Abstract (2013).
57. Ryan, J. F. *et al.* The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* **342**, 6164 (2013).
58. Kocot, K. M. *et al.* Phylogenomics reveals deep molluscan relationships. *Nature* **477**, 452–456 (2011).
59. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
60. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
61. Chenna, R. *et al.* Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**, 3497–3500 (2003).
62. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**, 403–405 (1998).
63. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **5**, 113 (2004).
64. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
65. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
66. Pitts, A. & Moroz, L. L. Computational workflow for analysis of gain and loss of genes in distantly related genomes. *BMC Bioinform.* **13** (suppl. 15) (2012).
67. Moroz, L. L. *et al.* Neuronal transcriptome of *Aplysia*: Neuronal compartments and circuitry. *Cell* **127**, 1453–1467 (2006).
68. Fuller, R. R., Moroz, L. L., Gillette, R. & Sweedler, J. V. Single neuron analysis by capillary electrophoresis with fluorescence spectroscopy. *Neuron* **20**, 173–181 (1998).
69. Lapainis, T., Rubakhin, S. S. & Sweedler, J. V. Capillary electrophoresis with electrospray ionization mass spectrometric detection for single-cell metabolomics. *Anal. Chem.* **81**, 5858–5864 (2009).
70. Nemes, P., Knolhoff, A. M., Rubakhin, S. S. & Sweedler, J. V. Metabolic differentiation of neuronal phenotypes by single-cell capillary electrophoresis-electrospray ionization-mass spectrometry. *Anal. Chem.* **83**, 6810–6817 (2011).



**Extended Data Figure 1 | Anatomy and neuromuscular systems of the ctenophore *Pleurobrachia bachei*.** **a–e**, Anatomy of the ctenophore, *Pleurobrachia bachei* A. Agassiz, 1860. Natural colouration of the major organs in live animal is shown. **a**, Details of the transparent *Pleurobrachia* body are shown including, **b**, the pharynx and tentacle sheaths (pockets). Eight rows of comb plates, called ctenes, are made of giant compound cilia that diffract light, creating iridescence. **c**, Comb rows in *Pleurobrachia* are constantly beating. The mouth and the aboral organ (AO) are located at the opposite poles of the animal (**a**, **c**). The AO controls complex coordinated behaviours of the animal. **d**, Ciliated furrows connect the AO and the ctenes to mediate behaviour. **e**, Tentacles have numerous contractile tentillae used to capture food with

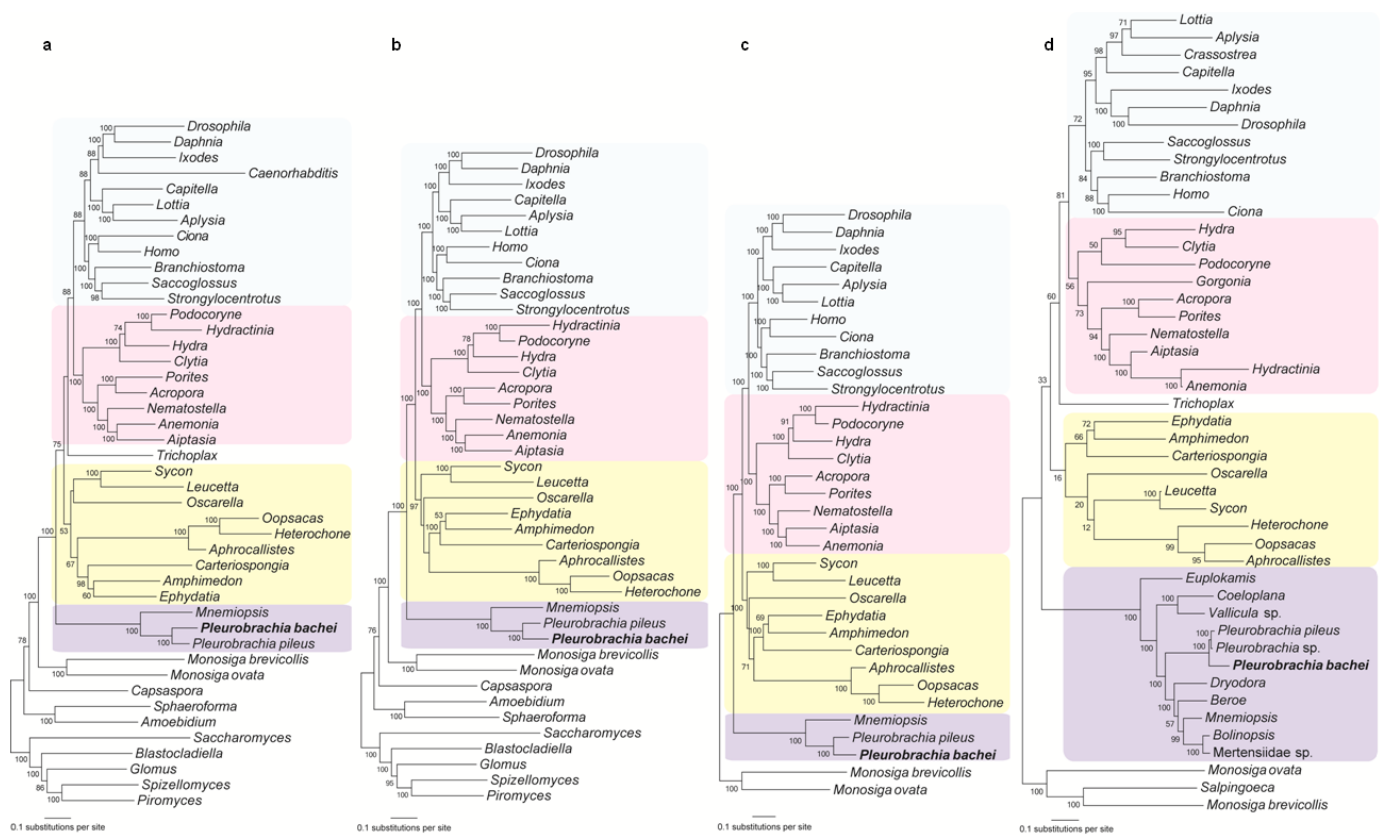
specialized glue cells or colloblasts (see also Fig. 1e). **f–h**, *Pleurobrachia* neural nets and muscles. **f**, Comb plate muscles (red) were revealed using *in situ* hybridization for  $\beta$ -tubulin and subepithelial neural net (green) revealed by tyrosinated  $\alpha$ -tubulin immunostaining. **g**, In this image comb cilia (green) were stained using tyrosinated  $\alpha$ -tubulin antibodies (green) whereas underlying comb plate muscles were visualized by phalloidin (a muscle marker) that did not stain neurons. **h**, Organization of the subepithelial neural net around the mouth as revealed by tyrosinated  $\alpha$ -tubulin antibodies (whole-mount preparation). Scale bars: 120  $\mu$ m (**f**); 100  $\mu$ m (**g**); 200  $\mu$ m (**h**). See Supplementary Methods 10 and 11.



**Extended Data Figure 2 | DNA methylation and active DNA demethylation in *Pleurobrachia bachei*.** CpG DNA methylation facilitates the elimination of CpG dinucleotides over evolutionary time<sup>66</sup>. **a**, Histogram shows relative occurrences of different dinucleotides in genomes of *P. bachei* (red bars), *Drosophila melanogaster* (green bars, no DNA methylation) and *Homo sapiens* (blue bars). The *P. bachei* genome contains 2.3% CpG dinucleotides, which is much lower than the expected random frequency and, therefore, indicative of a genome that undergoes methylation compared to humans<sup>66</sup>. **b**, DNMT genealogy tree. The enzyme DNA methyltransferase (DNMT), which catalyses transfer of a methyl group to DNA to form 5-methyl cytosine (5-mC), is present in *Pleurobrachia*. **c**, TET family of enzymes catalyses active DNA

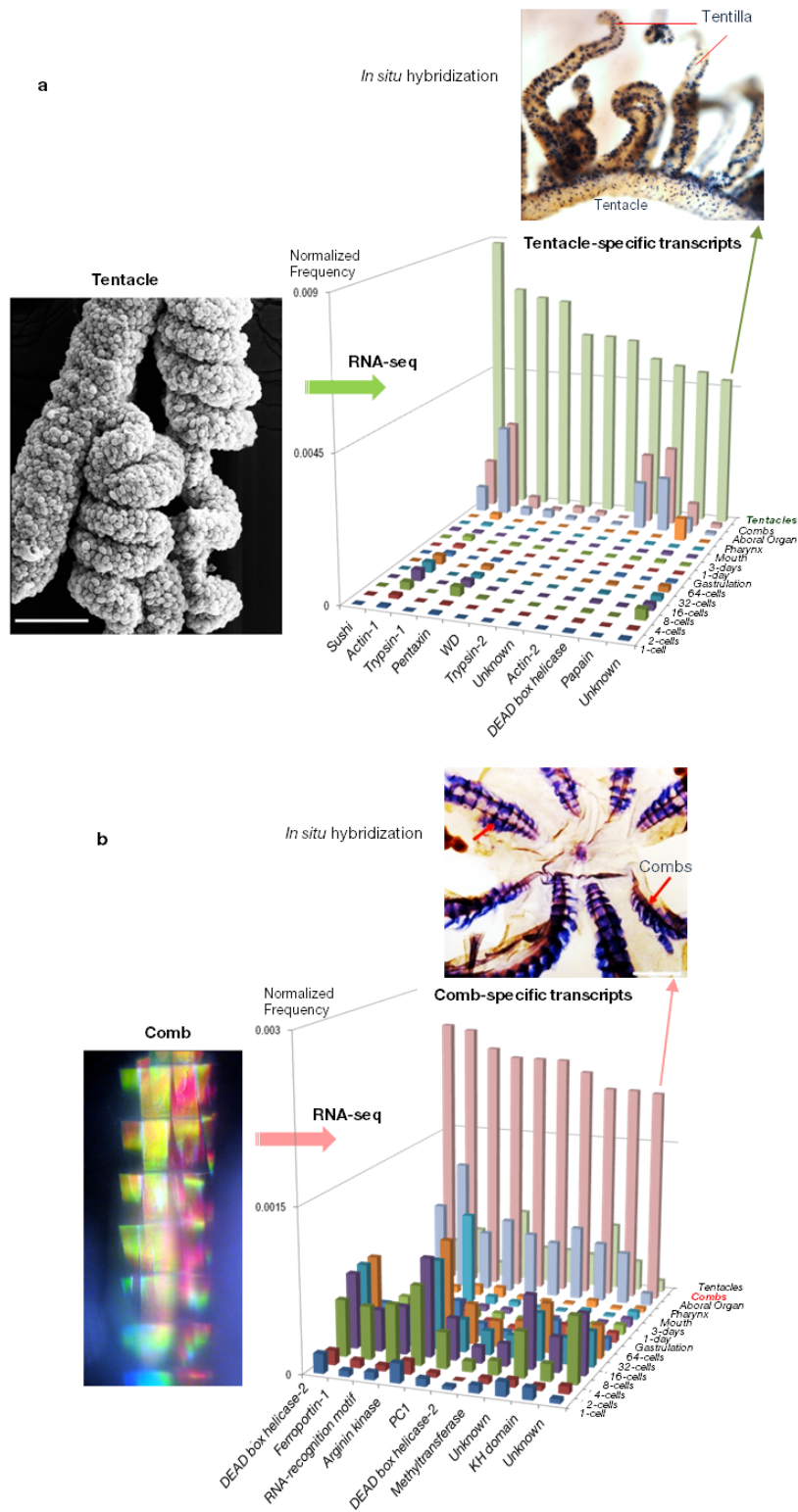
demethylation via formation of 5-hydroxymethyl cytosine (5-hmC, the 6th DNA base). RNA-seq profiling reveals differential expression for DNMT and TET-like genes during development and in adult *P. bachei*. Both DNMT and TET-like genes are predominantly expressed during cleavage starting from the 1st division. However, the TET-like gene is also highly expressed in adult combs (asterisk). The y axis shows a normalized expression level for each transcript. **d**, ELISA-based colorimetric assays validate the presence of both 5-mC and 5-hmC in the *P. bachei* genome (the rat brain is used as a positive control;  $n = 6$  for *Pleurobrachia* and  $n = 3$  for rat; data shown as mean  $\pm$  s.e.m., see Supplementary Methods 8 and Supplementary Data section 3 for details).





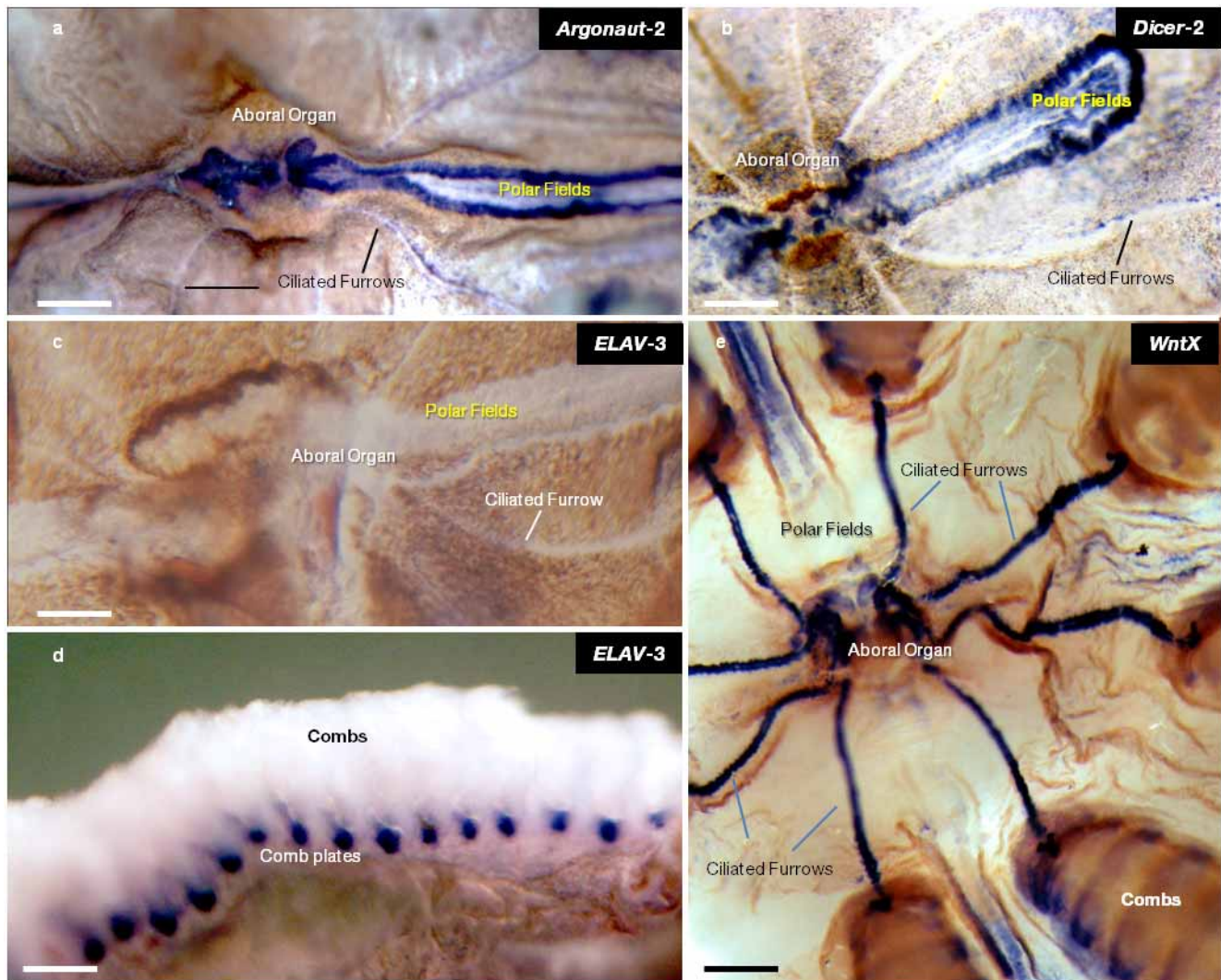
**Extended Data Figure 3 | Animal phylogeny.** **a**, Phylogeny of Metazoa based on 586 genes. Topology inferred using RAxML 7.2.7 and maximum likelihood with the CAT +WAG + F model with all taxa from Supplementary Table 12. Bootstrap support values are listed at each node. Colour coding: purple, ctenophore; yellow, Porifera; pink, Cnidaria; light blue, Bilateria. **b**, Removal of fast-evolving taxa *Trichoplax* and *Caenorhabditis* improves topological robustness. Topology inferred using RAxML 7.2.7 and maximum likelihood with the CAT +WAG + F model with all taxa from Supplementary Table 12 except *Trichoplax* and *Caenorhabditis*. Bootstrap support values are listed at

each node. **c**, Removal of distant out-groups such as Fungi and Filasterea further improves topological robustness. Topology inferred using RAxML 7.2.7 using maximum likelihood with the CAT +WAG + F model with all taxa from Supplementary Table 12 except *Trichoplax*, *Caenorhabditis* and non-choanoflagellate out-groups. Bootstrap support values are listed at each node. **d**, Analysis with improved ctenophore taxon sampling based on 114 genes. Topology inferred using RAxML 7.2.7 using maximum likelihood with the CAT +WAG + F model with all taxa from Supplementary Table 13. Bootstrap support values are listed at each node.



**Extended Data Figure 4 | Identification of tentacle- and comb-specific transcripts.** **a**, Identification of tentacle-specific transcripts. The left photo shows SEM of a *Pleurobrachia* tentacle with two branching tentillae densely covered with hundreds of colloblasts or glue cells. Comparative transcriptome (RNA-seq) profiling among major organs allowed us to identify several dozen genes differentially or uniquely expressed in tentacles. The histogram shows illustrative examples of some of these genes with a normalized expression level (y axis) for each represented transcript. One of these *Pleurobrachia*-specific genes we named tentillin (green arrow). *In situ* hybridization experiments ( $n = 9$ ) revealed a remarkable cell-specificity expression pattern for tentillin in all main tentacle branches and tentillae, possible labelling colloblasts or

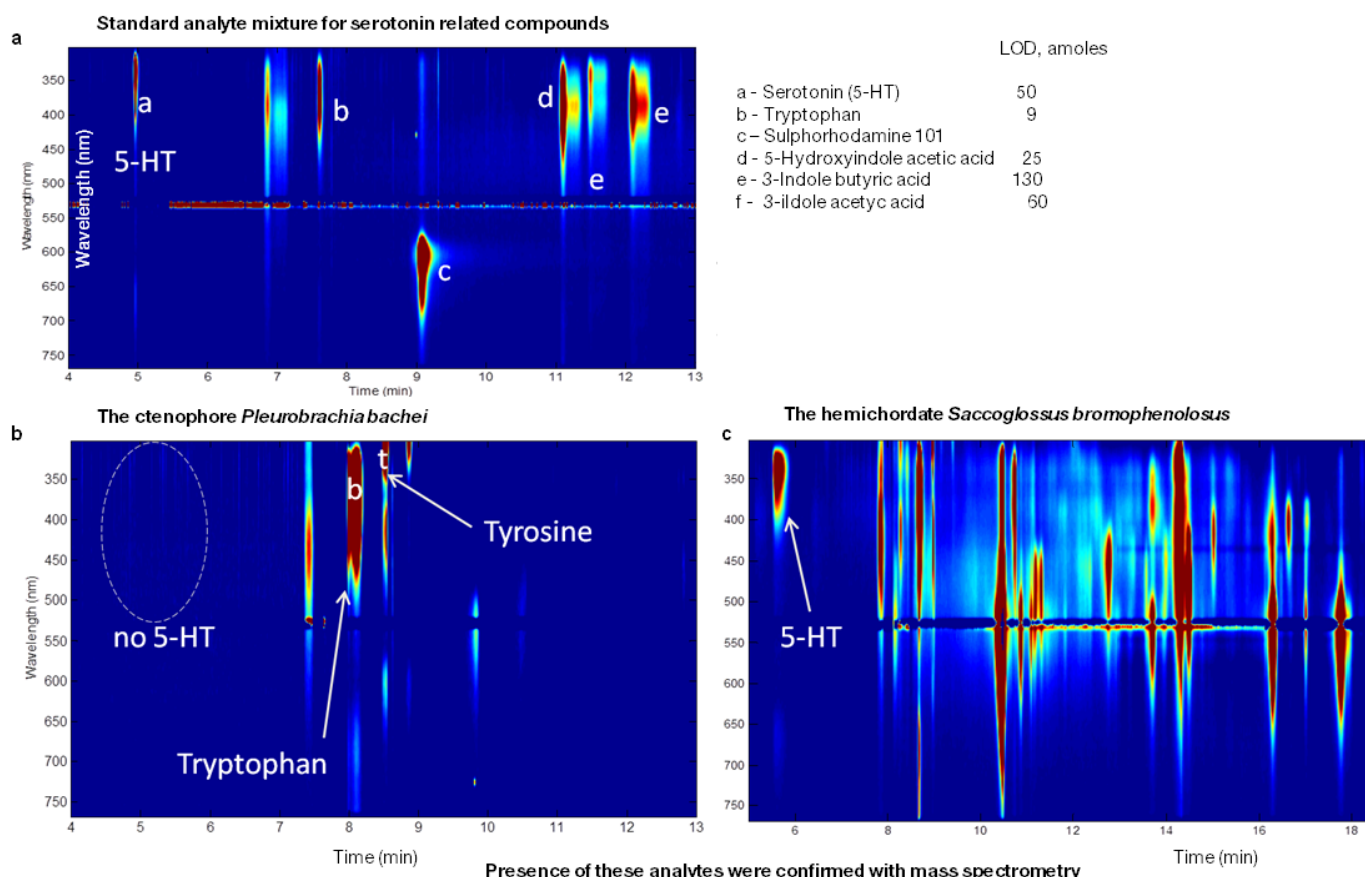
associated secretory cells. Scale bar: 50  $\mu$ m. **b**, Identification of comb-specific transcripts. The left photo shows a microscopic image of one comb row from an intact animal. The natural colouration is a reflection of the iridescence patterns produced from large cilia forming combs. Comparative transcriptome (RNA-seq) profiling among major organs allowed us to identify several hundreds of genes differentially or uniquely expressed in combs. The histogram shows illustrative examples of some of these genes with a normalized expression level (y axis) for each represented transcript (see Supplementary Methods 4.2.3.6, 4.2.3.7 and 10, all sequences used in the analysis can be found in Supplementary Tables 29, 30 and 32).



**Extended Data Figure 5 | Illustrative examples of tissue-specific gene expression in *Pleurobrachia bachei*.** **a, b,** *Dicer* and *Argonaut* are predominantly expressed in structures associated to sensory and integrative functions. These include the aboral organ, polar fields and combs. Note a relatively weak staining of other cell types in the skin and following ciliated furrows in *Dicer* and *Argonaut* preparations. **c, d,** *Pleurobrachia* ELAV is expressed in combs and not in neurons. ELAVs are RNA-binding proteins and they are considered as pan-neuronal markers (see Supplementary Data 5.6.1). However, in *Pleurobrachia* ELAV expression has not been detected in neural tissues or cells with recognizable neuronal-like appearances. *In situ* hybridization for *Pleurobrachia* ELAV3 (**c, d**) shows the highest levels of expression in the adult comb plate but not in any of the neural tissues or organs

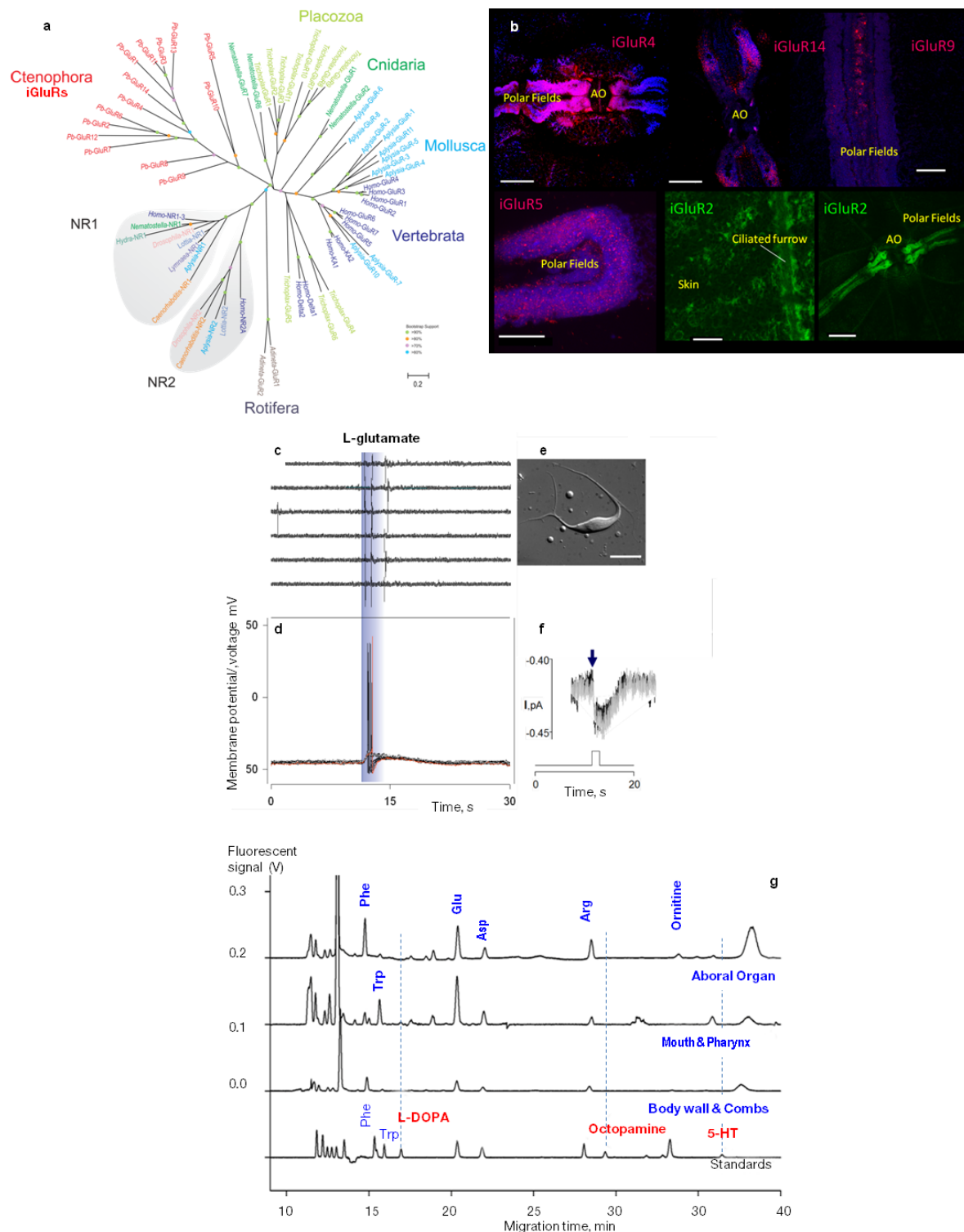
enriched with neurons such as the aboral organ and polar fields. **e,** *WntX* is selectively expressed in the aboral organ (AO) and major conductive pathways of *Pleurobrachia*, indicating its involvement in integrative and neural-like functions (*in situ* hybridization on a whole-mount preparation). One of the highest *WntX* expressions is found in AO and ciliated furrows whereas the polar fields showed a moderate expression level associated to their central regions. *In situ* hybridization was performed on whole mounts using DIG-labelled probes (see details in the Supplementary Methods, all *in situ* hybridization were performed at least on 4–5 different animals and these are representative photos for these experiments). Scale bars: 500  $\mu$ m (**a–d**); 800  $\mu$ m (**e**).





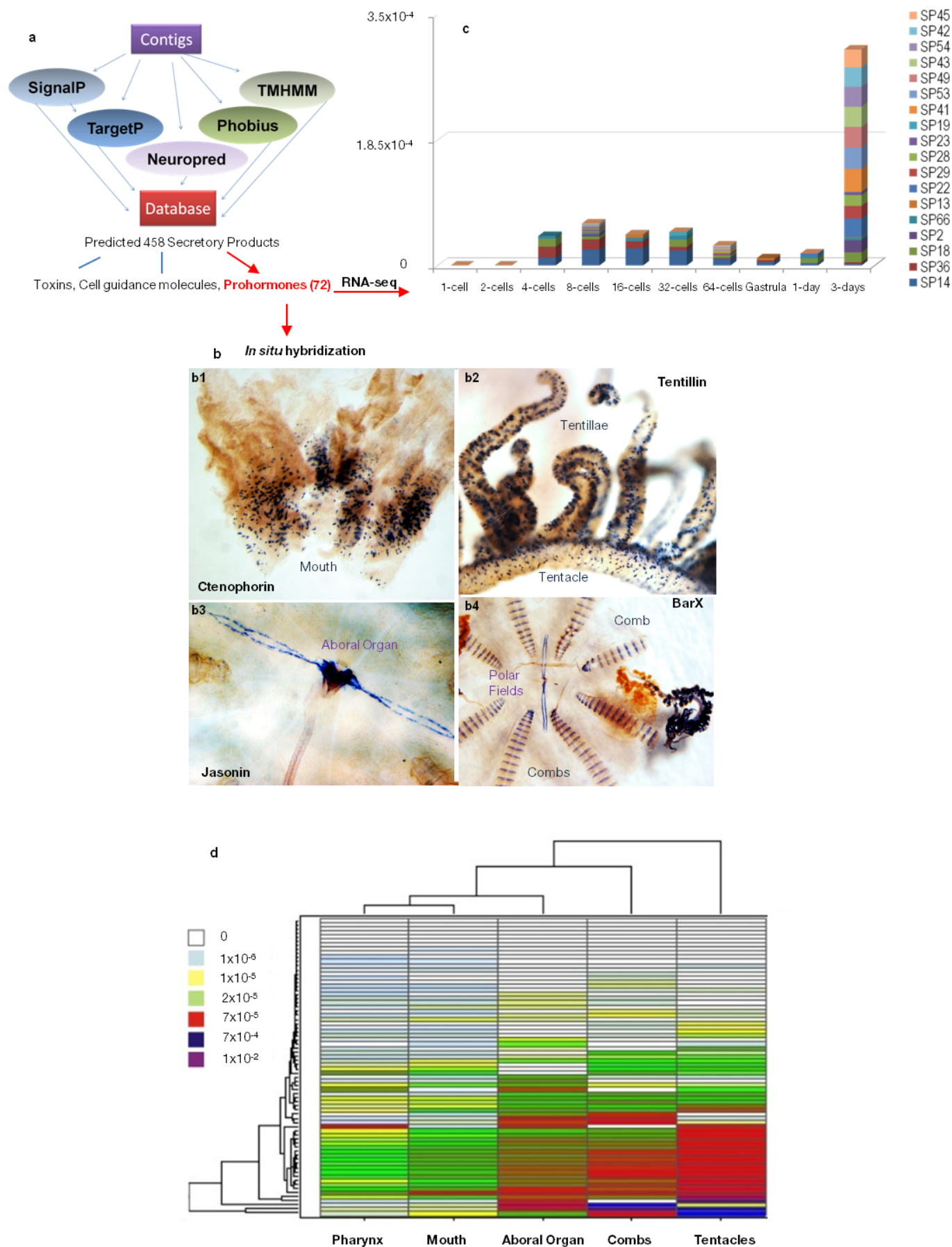
**Extended Data Figure 6 | Absence of serotonin in ctenophores.** Here we used nanolitre volume sampling, capillary electrophoresis separation and wavelength-resolved native fluorescence detection as described for ultra-sensitive assay of 5-hydroxytryptamine (serotonin or 5-HT) and related metabolites (a, the top electropherogram and the table with standards used).

Limits of detection (LODs) range from the low attomole to the femtomole range, with 5-HT LODs being approximately 20–50 attomoles. **b, c**, Using this assay we failed to detect 5-HT in *Pleurobrachia* (**b**,  $n = 6$ ) but 5-HT was reliably detected in the hemichordate *Saccoglossus* (**c**) and molluscs<sup>62</sup>. See details in Supplementary Methods 17 and Supplementary Table 22 for quantification.



**Extended Data Figure 7 | 1-glutamate as a transmitter candidate in *Pleurobrachia bachei*.** **a**, The ionotropic glutamate receptors (iGluRs) are diverse and underwent substantial adaptive radiation within the Ctenophora lineage. Phylogenetic analysis shows *Pleurobrachia* iGluRs share highest identity to each other forming a distinct branch on the tree topology (Supplementary Data 5.9). **b**, Differentially expressed iGluR subtypes in *Pleurobrachia bachei* (red and green labelling with fluorescent *in situ* hybridization protocols). Dark blue fluorescence is DAPI nuclear staining. AO, aboral organ. Scale bars: 100  $\mu$ m (top left and top middle); 60  $\mu$ m (top right); 50  $\mu$ m (bottom left); 30  $\mu$ m (bottom middle); 200  $\mu$ m (bottom right). **c–f**, Glutamate-induced action potentials and currents in muscle cells. **c**, Typical responses of ctenophore muscle cells to glutamate pulses recorded

extracellularly (as individual action potentials/contractions from a single muscle cell in response to local application of glutamate, 1 mM), and **d**, from the same cell in whole-cell current clamp mode with clear action potentials. **e**, Isolated muscle cell. Scale bar, 25  $\mu$ m. **f**, Glutamate-activated whole-cell currents recorded from the same cell (as in **c**). Time course of application is depicted by the diagram below the voltage signal. Two responses (inward current) are shown. The holding potential was  $-70$  mV (Supplementary Methods 13–16). **g**, Representative electropherograms show capillary electrophoresis separation with laser-induced fluorescence detection from different organs in *Pleurobrachia bachei* ( $n = 5$ ) for transmitter candidate identification. The bottom electropherograms are standards (Supplementary Methods 17 and Supplementary Tables 23–25 for quantification).

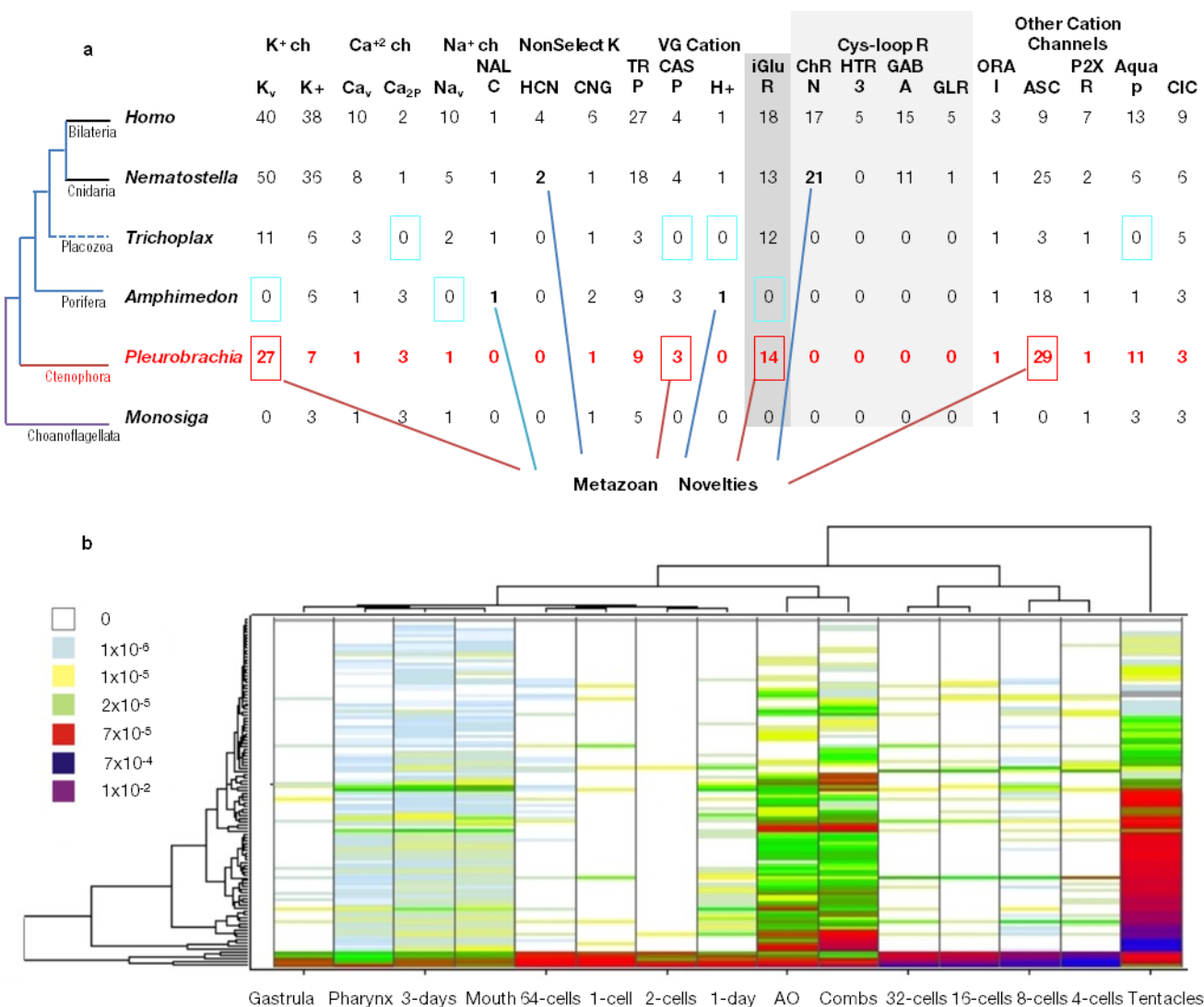




**Extended Data Figure 8 | Quest for putative secretory molecules in**

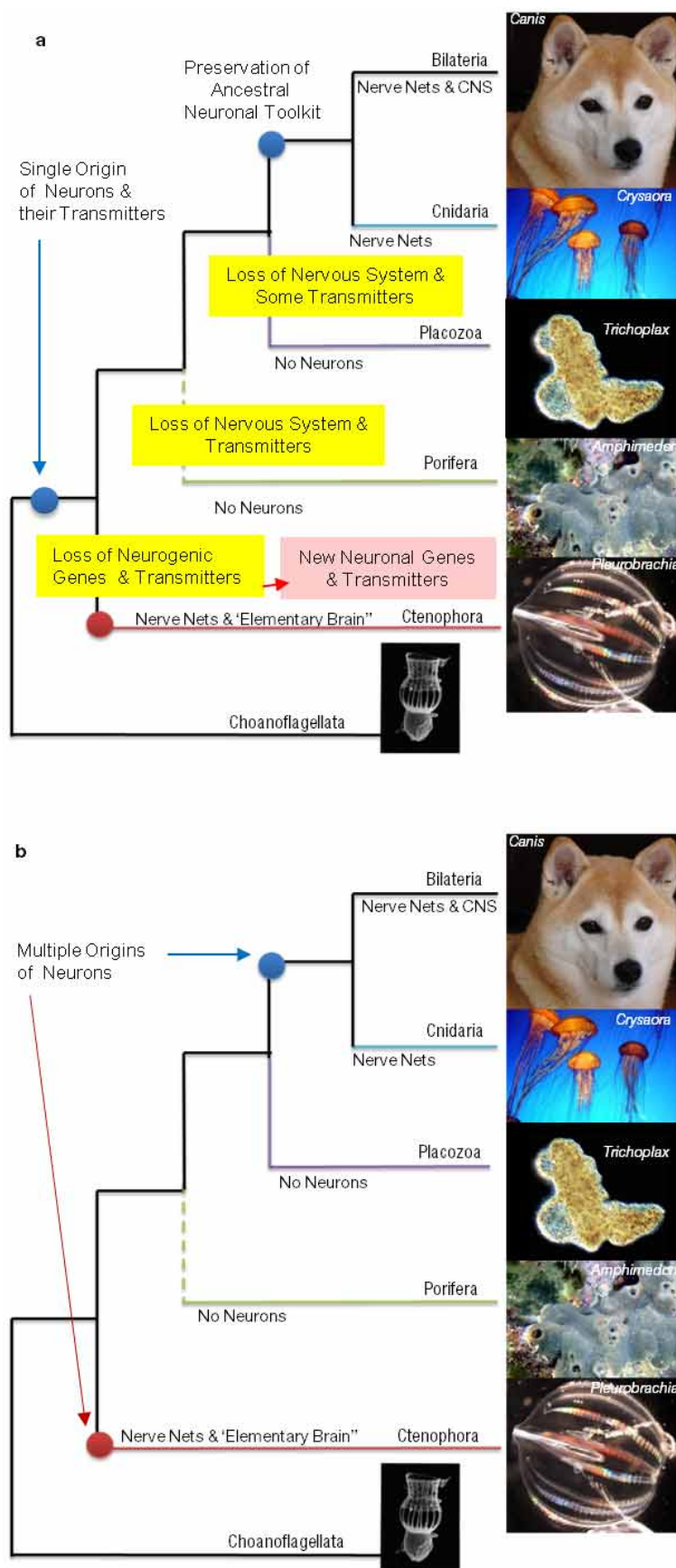
**ctenophores.** **a**, Computational pipeline for prediction of secretory products in *Pleurobrachia* and the overview of secretory products predicted from the *Pleurobrachia* gene models (Supplementary Method and Data sections 4.2.3.7 and 5.11, respectively). **b**, Expression of novel secretory molecules in ctenophores (DIG-labelled *in situ* hybridization, Supplementary Data 5.11 and Supplementary Methods 10). Each of the predicted secretory prohormone was selected based on its unique and/or highly differential expression pattern as revealed by RNA-seq profiling. Ctenophorin is uniquely expressed in polarized cells around the mouth of *Pleurobrachia* and we found its homologues in all ctenophore species we sequenced. Tentillin is a *Pleurobrachia*-specific gene, which is uniquely expressed in polarized secretory-like cells in tentillae and tentacles. Jansonin's expression is primarily restricted to polarized cells located

in the aboral organ and polar fields. For comparison, we showed a different but also cell-specific expression pattern of BarX transcription factor in cells of unknown identity localized in polar fields, comb plates and tentacles (see b4). **c, d**, Most predicted secretory products are expressed later in development and in adult organs of *Pleurobrachia* (RNA-seq). **c**, Expression patterns of 72 predicted prohormones in *P. bachei* indicates that 20 of them are present and differentially expressed in development (Supplementary Table 32 for all *Pleurobrachia* precursor sequences). Surprisingly 5 of these precursor mRNAs were found starting from the second cleavage stage whereas the rest are predominantly expressed on day 3 of development. This correlates to the first appearance of neurons in *Pleurobrachia* cydippid larva (see Supplementary Data 5.11 and Supplementary Method section 4.2.3.6 for the RNA-seq analysis).



**Extended Data Figure 9 | Diversity of ion channels in *Pleurobrachia bachei*.**  
**a**, Metazoan ion channel complement. The 112 ion channels identified in the *Pleurobrachia* genome are classified as voltage gated (v) or other gating such as second messengers. Receptor channels (R) are ligand-gated or ionotropic (iGluR, ChRN, HTR3, GABA and CLR) and indicated in grey. Metazoan novelties indicate type of ion channels absent in the choanoflagellates, the sister group to all animals. Coloured squares show channels: (1) primarily absent in ctenophores (pink); (2) secondarily lost in sponges or placozoans (dark yellow); (3) eumetazoan (Cnidaria+Bilateria) innovations (blue); or (4) examples of expansion of certain class of channels in some animal lineages (red). All

*Pleurobrachia* sequences used in the analysis can be found in Supplementary Table 31. **b**, Ion channels are predominantly expressed in tentacles, combs and aboral organ. Hierarchical clustering of 112 identified ion channels in developmental stages and adult tissues of *Pleurobrachia*. Adult organs involved in food capture and ciliated locomotion and integrative functions show significantly higher diversity and overall higher level of expression levels for most of ion channel types. Mobile tentacles had the highest expression of voltage-gated channels, in particular Ca<sub>v</sub> and Na<sub>v</sub>. The legend shows relative expression levels based on RNA-seq data (see Supplementary Methods 4.2.3.6).



**Extended Data Figure 10 | Two alternative scenarios of neuronal evolution.**  
**a**, Single origin of the neural system (monophyly) with possible loss of some neural molecular components in ctenophores as well as the possible secondary

loss of the entire nervous systems in sponges and placozoans. **b**, Multiple origins of neurons in animals as introduced and supported in this study (see main text discussion section for details).



# miR-34/449 miRNAs are required for motile ciliogenesis by repressing *cp110*

Rui Song<sup>1\*</sup>, Peter Walentek<sup>2\*</sup>, Nicole Sponer<sup>1\*</sup>, Alexander Klimke<sup>3</sup>, Joon Sub Lee<sup>1</sup>, Gary Dixon<sup>1</sup>, Richard Harland<sup>2</sup>, Ying Wan<sup>4</sup>, Polina Lishko<sup>1</sup>, Muriel Lize<sup>5</sup>, Michael Kessel<sup>3</sup> & Lin He<sup>1</sup>

**The *mir-34/449* family consists of six homologous miRNAs at three genomic loci. Redundancy of miR-34/449 miRNAs and their dominant expression in multiciliated epithelia suggest a functional significance in ciliogenesis. Here we report that mice deficient for all miR-34/449 miRNAs exhibited postnatal mortality, infertility and strong respiratory dysfunction caused by defective mucociliary clearance. In both mouse and *Xenopus*, miR-34/449-deficient multiciliated cells (MCCs) exhibited a significant decrease in cilia length and number, due to defective basal body maturation and apical docking. The effect of miR-34/449 on ciliogenesis was mediated, at least in part, by post-transcriptional repression of *Cp110*, a centriolar protein suppressing cilia assembly. Consistent with this, *cp110* knockdown in miR-34/449-deficient MCCs restored ciliogenesis by rescuing basal body maturation and docking. Altogether, our findings elucidate conserved cellular and molecular mechanisms through which miR-34/449 regulate motile ciliogenesis.**

MicroRNAs (miRNAs) encode a class of small, non-coding RNAs that regulate gene expression through post-transcriptional repression<sup>1–4</sup>. Although the initial discovery of miRNAs was made through classic forward genetics in worm development<sup>5,6</sup>, loss-of-function studies on most individual miRNAs yield no overt developmental defects in multiple organisms, suggesting strong functional redundancy among homologous miRNAs<sup>7,8</sup>. Redundant miRNAs can be generated from multiple genomic loci or transcribed from a single polycistronic precursor. Collectively, these redundant miRNAs could constitute the majority of expressed miRNAs in specific cell types<sup>9,10</sup>. Such extensive homology and dominant cell-type specific expression of a single miRNA family could confer a robust functional readout that can only be revealed by complete removal of redundant miRNAs.

*mir-34/449* miRNAs constitute a conserved family in vertebrates<sup>11–13</sup>, comprising three genomic loci, *mir-34a*, *mir-34b/34c* and *mir-449c/449b/449a* (*mir-449*), which encode six homologous miRNAs (miR-34a, 34b, 34c, 449a, 449b and 449c)<sup>14,15</sup> (Fig. 1a and Extended Data Fig. 1a). Sequence homology among miR-34/449 miRNAs, particularly at the seed region, predicts robust functional redundancy. miR-34/449 miRNAs are highly enriched in mucociliary epithelia that contain motile cilia<sup>10</sup>, which beat coordinately to generate fluid movement<sup>16,17</sup>. Structural and functional defects in motile cilia are associated with a human syndrome, primary cilia dyskinesia (PCD)<sup>16,17</sup>. Here we demonstrate that miR-34/449-deficient mice developed PCD-like respiratory and fertility phenotypes. Consistently, miR-34/449 deficiency in mice and frogs disrupts ciliogenesis in mucociliary epithelia, causing reduced cilia length and number due to impaired basal body maturation and apical docking. This is, at least in part, mediated by direct miR-34/449 repression of *Cp110*, a centriolar protein suppressing cilia assembly<sup>18,19</sup>. These findings reveal conserved cellular and molecular mechanisms underlying the functions of miR-34/449 in MCC ciliogenesis.

## PCD-like phenotype in *mir-34/449* TKO mice

To characterize *mir-34/449* functions, we generated triple knockout (TKO) mice deficient for all *mir-34/449* loci (*mir-34a*, *mir-34b/34c* and

*mir-449*)<sup>20</sup> (Extended Data Fig. 1b, c). Although *mir-449* resides in intron 2 of *cdc20b*<sup>21</sup>, *mir-449* deletion in mice does not negatively affect *cdc20b* expression (data not shown). TKO mice were born at a Mendelian ratio with normal body weight (Fig. 1b and Extended Data Fig. 1d); yet exhibited frequent postnatal mortality with only ~40% surviving to adulthood (Fig. 1b). TKO mice also exhibited growth attenuation with ~50% lower body weight than littermate-controlled double knockout (DKO) mice (*mir-34a*<sup>−/−</sup>; *mir-34b/34c*<sup>−/−</sup> or *mir-34a*<sup>−/−</sup>; *mir-449*<sup>−/−</sup>, Fig. 1c and Extended Data Fig. 1d).

Surviving TKO mice showed severe respiratory distress characterized by frequent coughing and sneezing (Extended Data Fig. 1e and Supplementary Video 1). Dying and surviving TKO mice displayed respiratory dysfunction, with excessive mucus accumulation in the paranasal cavities and increased susceptibility to respiratory infections (Fig. 1d and Extended Data Fig. 1f). Littermate-controlled *mir-34a*<sup>−/−</sup>; *mir-34b/34c*<sup>−/−</sup> or *mir-34a*<sup>−/−</sup>; *mir-449*<sup>−/−</sup> DKO mice phenotypically resembled wild-type mice, without obvious developmental or respiratory defects (Fig. 1d and Extended Data Fig. 1f, g).

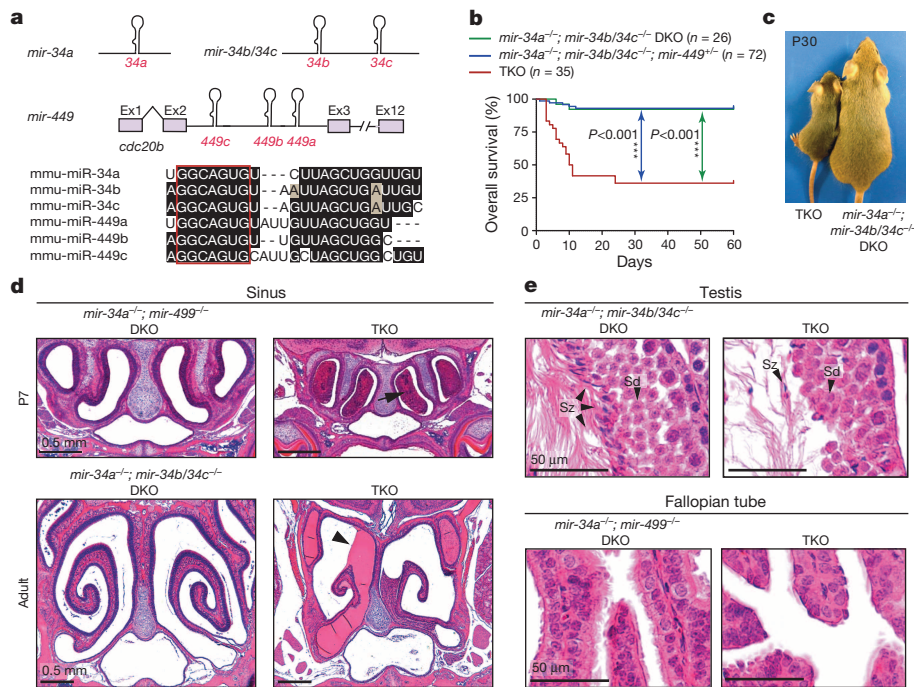
Unlike phenotypically normal DKO controls (*mir-34a*<sup>−/−</sup>; *mir-449*<sup>−/−</sup> or *mir-34a*<sup>−/−</sup>; *mir-34b/34c*<sup>−/−</sup>), surviving TKO males and females were infertile, generating no pregnancies when mated with wild-type animals. TKO males exhibited defective spermatogenesis during differentiation from elongating spermatids to spermatozoa (Fig. 1e and Extended Data Fig. 2a, b), when flagella formation occurs<sup>17</sup>. TKO females exhibited a decrease in epithelial ciliation of the fallopian tube, presumably causing defects in oocyte transport<sup>17</sup> (Fig. 1e and Extended Data Fig. 2a, c). Altogether, the TKO mouse phenotype resembled symptoms of a subset of PCD patients, exhibiting predominant respiratory and fertility defects without hydrocephaly or left-right asymmetry defects<sup>22,23</sup> (Extended Data Fig. 2d).

## Ciliogenesis defects in *mir-34/449* TKO mice

Mature miR-34/449 miRNAs were enriched in organs containing motile cilia, including lung, brain, testis and female reproductive organs (Extended Data Fig. 3a). We specifically detected and quantified individual miR-34/449

<sup>1</sup>Division of Cellular and Developmental Biology, MCB Department, University of California at Berkeley, Berkeley, California 94705, USA. <sup>2</sup>Division of Genetics, Genomics and Development, Centre for Integrative Genomics, MCB Department, University of California at Berkeley, Berkeley, California 94705, USA. <sup>3</sup>Department of Molecular Cell Biology, Max Planck Institute for Biophysical Chemistry, Goettingen 37077, Germany. <sup>4</sup>The Third Military Medical University, Chongqing 400038, China. <sup>5</sup>Department of Molecular Oncology, University of Goettingen, Goettingen 37073, Germany.

\*These authors contributed equally to this work.



**Figure 1 | *mir-34/449* TKO mice exhibit defective mucociliary airway clearance and infertility.** **a**, Gene structure (top) and sequence alignment (bottom) of mouse miR-34/449 miRNAs. Red box indicates seed sequences. **b**, TKO mice exhibit frequent postnatal mortality shown as log-rank test. **c**, Surviving TKO mice display postnatal growth retardation. **d**, Excessive mucus accumulation and infection in paranasal sinuses of dying TKO at P7

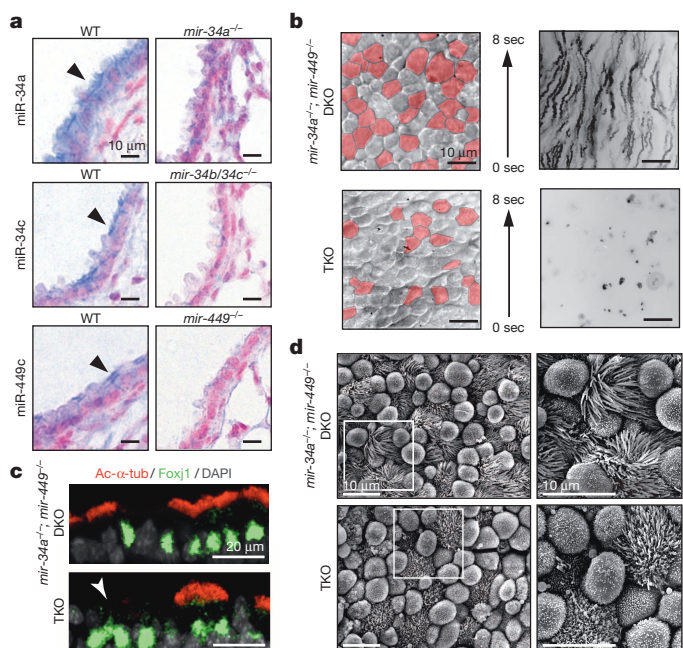
(top) and surviving adult TKO mice (bottom). Arrow, infection; arrowhead, mucus accumulation; n = 15. **e**, Adult TKO males and females are infertile. Although early spermatids (Sd) are developed, few intact spermatozoa (Sz) are generated (top, n = 3). A significant MCC reduction is observed in TKO fallopian tubes (bottom, n = 3).

miRNAs using single knockout and TKO controls (Fig. 2a and Extended Data Fig. 3b, c). *In situ* hybridization revealed high-level miR-34/449 expression in respiratory epithelia, with miR-34a being expressed broadly in multiple cell types, and miR-34c or miR-449c being enriched specifically in airway MCCs (Fig. 2a and Extended Data Fig. 3d).

A major symptom of PCD is dysfunctional airway clearance<sup>16,17</sup>. Defective mucociliary clearance in *mir-34/449* TKO mice, along with the MCC-specific miR-34/449 expression, prompted us to examine the roles of miR-34/449 in airway MCCs. High-speed imaging revealed a slow and limited fluid movement in TKO tracheal explants, accompanied by a significant reduction of visibly ciliated MCCs (Fig. 2b and Supplementary Video 2). This contrasts the effective anteriorward fluid flow in wild-type and DKO tracheal explants (Fig. 2b and Supplementary Videos 2 and 3).

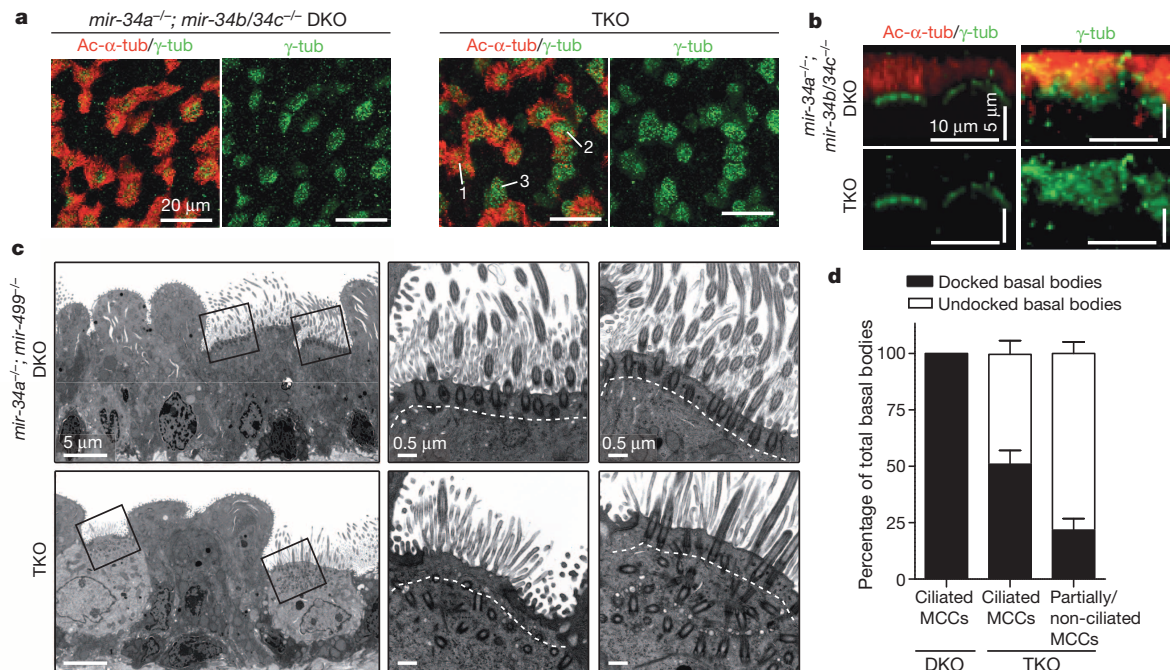
The decrease of visible MCCs in TKO tracheas could reflect defective cell fate specification or ciliogenesis. We analysed *mir-34a*<sup>-/-</sup>; *mir-449*<sup>-/-</sup> DKO and TKO tracheas for *Foxj1*, a master regulator of motile ciliogenesis<sup>24</sup>, and acetylated  $\alpha$ -tubulin (Ac- $\alpha$ -tub), a cilia marker. Both DKO and TKO tracheas had *Foxj1* immunofluorescence staining and *Foxj1* mRNA levels comparable to wild-type controls (Fig. 2c and Extended Data Fig. 3e). Nevertheless, a portion of *Foxj1*-positive cells lacked cilia in TKO tracheas, yet most *Foxj1*-positive cells were fully ciliated in DKO and wild-type tracheas (Fig. 2c, and data not shown). This suggests normal cell fate specification with defective ciliation in *mir-34/449*-deficient MCCs. Scanning electron microscopy supported this finding, revealing a significant reduction in cilia length and number per MCC in TKO, but not in DKO and wild-type tracheal epithelia (Fig. 2d and Extended Data Fig. 3f). Notably, TKO MCCs displayed a spectrum of ciliation phenotypes (non-ciliated, partially or fully ciliated; Fig. 2d), possibly due to the mixed genetic background.

MCC ciliogenesis is characterized by the multiplication of basal bodies, which, after docking to the apical membrane, act as microtubule-organizing centres to assemble motile axonemes<sup>25,26</sup>. We stained TKO tracheas with antibodies against Ac- $\alpha$ -tub and  $\gamma$ -tubulin ( $\gamma$ -tub) to visualize



**Figure 2 | miR-34/449 deficiency causes ciliogenesis defects in respiratory MCCs.** **a**, miR-34/449 are strongly enriched in MCCs of respiratory epithelia (arrowheads), shown by *in situ* hybridization (n = 2). **b**, TKO tracheal epithelia exhibit defective mucociliary clearance demonstrated by live imaging of fluorescent bead transport. Red, visibly ciliated MCCs; n = 4. **c**, Cell-fate specification of MCCs is unaffected in TKO tracheas. Immunofluorescence staining for *Foxj1* (a MCC marker) is unaltered in TKO tracheas, yet a large number of *Foxj1*-positive MCCs (arrowhead) have decreased staining for Ac- $\alpha$ -tub (a cilia marker). n = 3. **d**, TKO tracheal MCCs have a significant reduction in cilia number and length, revealed by scanning electron microscopy (n = 3).





**Figure 3 | miR-34/449 deficiency causes defective basal body docking in mouse airway MCCs.** **a**, TKO tracheas exhibit ciliation defects, shown by immunofluorescence staining for Ac- $\alpha$ -tub (cilia) and  $\gamma$ -tubulin (basal bodies). Numbers on the image indicate fully (1), partially (2) or non-ciliated (3) MCCs;  $n = 3$ . **b–d**, Basal bodies fail to dock to the apical membrane of TKO MCCs.

cilia and basal bodies of MCCs, respectively. Consistent with previous observations, Ac- $\alpha$ -tub staining was greatly decreased in TKO tracheas, yet the percentage of  $\gamma$ -tub-enriched MCCs remained normal (Fig. 3a and Extended Data Fig. 4a, b). Air–liquid interface (ALI) culture of primary TKO tracheal epithelia yielded a similar observation (Extended Data Fig. 4c). Thus, basal body multiplication occurred normally in miR-34/449-deficient MCCs following MCC cell fate specification, yet cilia formation was impaired.

### Impaired basal body docking in *mir-34/449* TKO MCCs

Basal body docking to the apical MCC membrane is essential for proper ciliogenesis<sup>25,26</sup>. In *mir-34a<sup>-/-</sup>; mir-34b/34c<sup>-/-</sup>* DKO MCCs,  $\gamma$ -tub staining was apically localized, indicating normal basal body docking (Fig. 3b). In contrast,  $\gamma$ -tub staining was diffuse in TKO tracheal MCCs and ALI culture, suggesting defective basal body docking to or stabilization at the apical membrane (Fig. 3b and Extended Data Fig. 4d). Transmission electron microscopy revealed well-aligned basal bodies at the apical membrane of wild-type and DKO MCCs (Extended Data Fig. 4e). Yet in TKO MCCs, a significant percentage of basal bodies were mislocalized to the cytoplasm and unable to grow cilia, and those apically docked generally formed shorter cilia (Fig. 3c, d). The extent of ciliation defects correlated well with basal body docking defects, suggesting aberrant basal body docking as a key mechanism for impaired ciliation (Fig. 3d). Defective ciliation and basal body docking in TKO MCCs also correlated with a disturbed apical actin organization (Extended Data Fig. 4f).

Despite defective basal body docking in TKO MCCs, the structural components of basal bodies, either apically docked or mislocalized, remained largely intact (Extended Data Fig. 5a, b). Although axoneme structure was unaffected in TKO MCCs, basal body orientation and ciliary axoneme directionality exhibited mild defects (Extended Data Fig. 5c, d), which, in combination with the ciliation defects, probably evoked a strong mucociliary clearance phenotype.

### miR-34/449 functions in *Xenopus* MCCs

Mammalian and *Xenopus* miR-34/449 miRNAs are not only conserved in sequence<sup>14,15</sup>, but also in MCC-specific expression and ciliogenesis<sup>10</sup>.

**b**, Lateral projections of confocal micrographs shown in **a**. **c**, Transmission electron microscopy (TEM) confirms basal body docking defects in TKO MCCs ( $n = 3$ ). **d**, Quantification of basal body docking based on TEM studies in **c**. Docked and undocked basal bodies exhibit a distance  $\leq 0.3 \mu\text{m}$  and  $> 0.3 \mu\text{m}$  to the apical membrane, respectively. Error bar, s.e.m.

As a mucociliary epithelium, the *Xenopus* embryonic epidermis resembles mammalian airway epithelia in MCC development and function<sup>27</sup>. As in mouse, knockdown of *Xenopus* miR-34a/34b/449a by morpholino (MO) injection (miR-34/449 MOs) reduced cilia number and length in tadpole epidermal MCCs (Fig. 4a, b). No obvious defects in embryonic development, hydrocephalus, MCC cell fate specification or other cell type specification were observed (Extended Data Fig. 6a–f). In miR-34/449 morphants, a significant portion of MCCs were either partially ciliated ( $> 50\%$ ), or devoid of cilia ( $29\% \pm 17\%$ ), with frequent, unorganized subapical Ac- $\alpha$ -tub enrichment indicating defective basal body docking<sup>28</sup> (Fig. 4b). Consistently, basal bodies detected by  $\gamma$ -tub or Sas6-GFP<sup>29</sup> exhibited irregular distribution and frequently failed to form cilia in miR-34/449-deficient MCCs (Fig. 4c and Extended Data Fig. 6g).

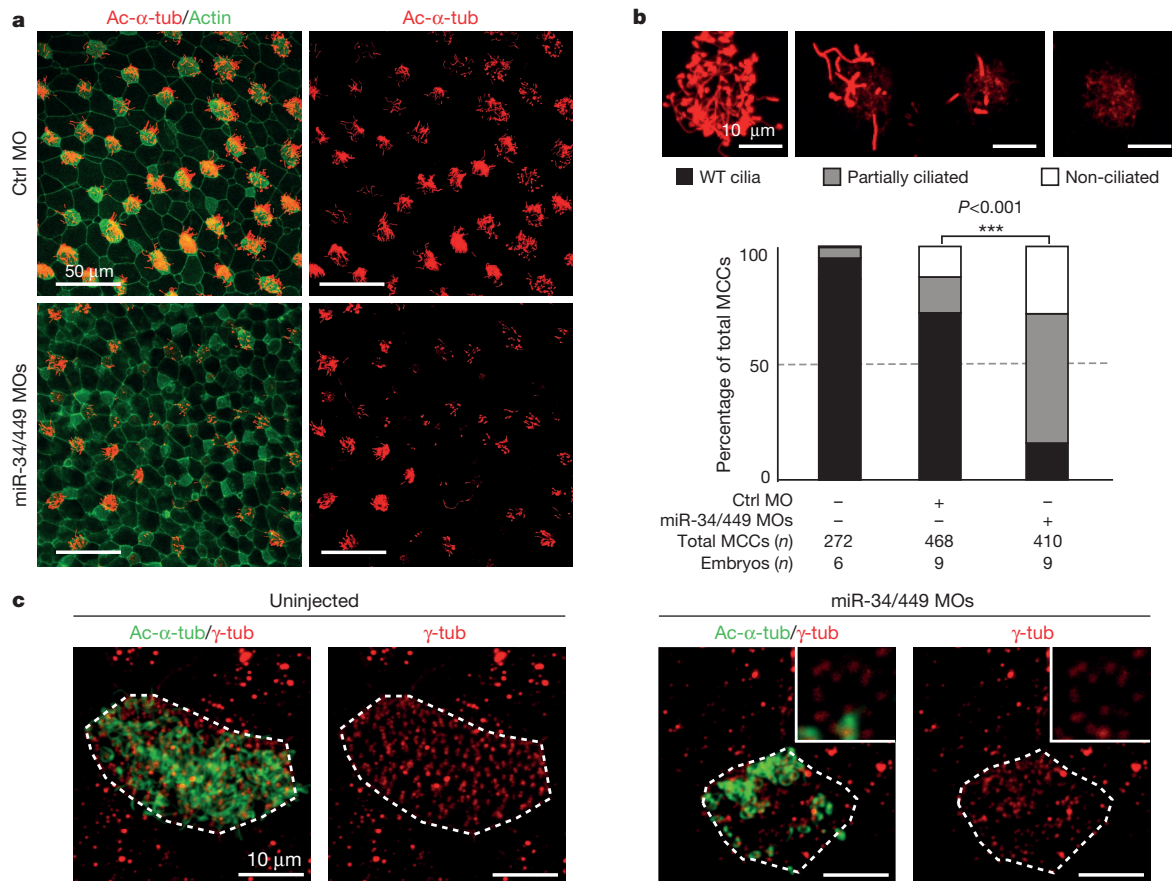
### miR-34/449 miRNAs directly repress *cp110* in MCCs

The miR-34/449 increase during MCC differentiation<sup>30</sup> predicts a decrease of functionally important targets, which invariably contain miR-34/449 binding site(s) in the 3' untranslated region (3' UTR)<sup>31,32</sup>. We analysed published gene expression profiles of tracheal MCCs during differentiation<sup>30,33</sup>, then selected 57 potential miR-34/449 targets for quantitative polymerase chain reaction with reverse transcription (qRT-PCR) validation in DKO and TKO tracheas, and finally narrowed down to those with important ciliogenesis functions (Extended Data Table 1). *Cp110* emerged as a strong candidate, containing two miR-34/449 binding sites (Extended Data Fig. 7a) and exhibiting miR-34/449-dependent repression *in vivo* and in luciferase assays (Fig. 5a, b and Extended Data Fig. 7b, c).

Cp110 is a distal centriolar protein suppressing primary cilia assembly<sup>18,19</sup>. Aberrant Cp110 retention in mother centrioles is correlated with impaired basal body docking<sup>34</sup>. In addition to its well-characterized roles in primary cilia, Cp110 is also implicated as an important regulator of motile ciliogenesis<sup>35,36</sup>.

*mir-34/449* and *cp110* are conserved in mice and frogs. *Xenopus cp110* contains one predicted miR-34/449 binding site (Extended Data Fig. 7d), suggesting a selective pressure to preserve miR-34/449–*cp110* regulation. miR-34/449 and *cp110* levels are inversely correlated during MCC





**Figure 4 | miR-34/449 deficiency causes defective ciliogenesis in the *Xenopus* embryonic epidermis.** **a**, MCCs in miR-34/449 morphants show reduced cilia length and number, demonstrated by immunofluorescence for Ac- $\alpha$ -tub (cilia) and phalloidin-488 (actin). **b**, Quantification of MCC ciliation in **a**.  $P < 0.001$  from a  $\chi^2$ -test. **c**, Co-staining of Ac- $\alpha$ -tub (cilia) and  $\gamma$ -tub (basal

bodies) in miR-34/449 morphants reveals uneven/aggregated distribution of basal bodies, which frequently fail to form cilia. Total numbers of embryos/total number of cells analysed were uninjected (4/14; a total of 14 cells from a total of 4 injected embryos) and miR-34/449 MOs (5/30). Embryos were derived from at least two females and independent fertilizations per experiment.

differentiation in *Xenopus*; and *cp110* mRNA was derepressed in miR-34/449 morphants (Extended Data Fig. 7e, f, g). Strikingly, *cp110* knockdown in miR-34/449-deficient MCCs significantly rescued ciliation defects (Fig. 5c and Extended Data Fig. 8a, b).

We subsequently examined, in control and miR-34/449 morphants, the localization of Centrin4 (ref. 37), an important basal body component whose loss-of-function causes a PCD-like phenotype<sup>38</sup>. Although strong Centrin4 foci were enriched apically in controls, these foci significantly decreased in intensity and failed to localize to the apical membrane of miR-34/449-deficient MCCs (Fig. 5d and Extended Data Fig. 8c). These findings demonstrated defective basal body docking and decreased basal body incorporation of Centrin4 in miR-34/449-deficient MCCs, both of which were rescued by co-injection of *cp110* MO (Fig. 5d).

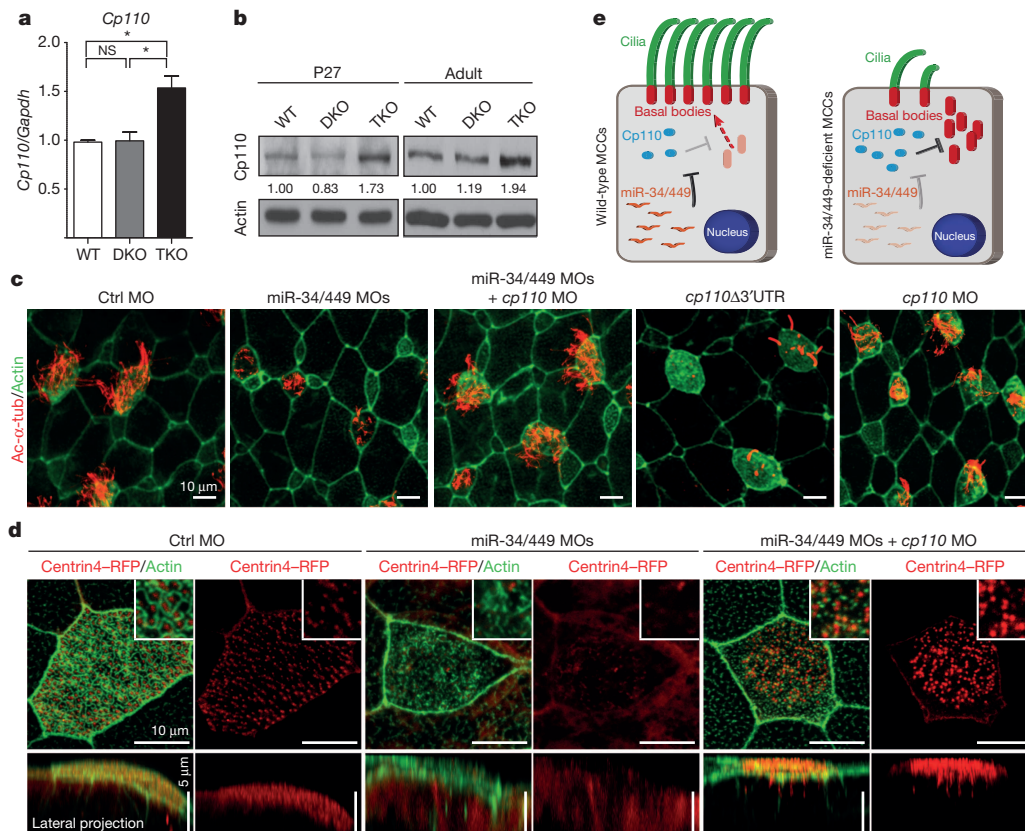
Consistently, *cp110* overexpression generally phenocopied miR-34/449 knockdown, causing impaired ciliation and decreased Centrin4 incorporation into basal bodies, without affecting MCC cell fate specification or apical actin organization (Fig. 5c and Extended Data Fig. 8b, d, e and 9a). *cp110* with a 3' UTR deletion exhibited a stronger phenotype than full-length *cp110*, suggesting *cp110* repression by miR-34/449 even during overexpression (Extended Data Fig. 7d and 8b, d). Surprisingly, *cp110* MO injection alone also gave rise to reduced cilia number and length, and aberrant basal body aggregation in MCCs (Fig. 5c and Extended Data Fig. 8a, b and 9b). Thus, miR-34/449 miRNAs mediate precise Cp110 regulation during vertebrate MCC ciliogenesis (Fig. 5e).

## Discussion

miR-34/449 represent the first non-coding RNAs whose deficiency causes a PCD-like airway and fertility phenotype. miR-34 miRNAs have

been mostly characterized as p53 targets that elicit multiple tumour suppressor effects<sup>20,39</sup>; yet the MCC-specific miR-34/449 expression and functions are likely p53-independent (data not shown). Although most characterized PCD mutations affect structural components of the ciliary axoneme<sup>17,40</sup> or basal body structure<sup>41</sup>, miR-34/449 miRNAs regulate ciliogenesis by promoting basal body maturation and docking without affecting overall basal body structure. Interestingly, redundant miR-34/449 miRNAs are not functionally equivalent in mice. One *mir-34b/34c* or *mir-449* allele is sufficient for proper MCC ciliation; two intact *mir-34a* alleles in *mir-34b/34c*<sup>-/-</sup>; *mir-449*<sup>-/-</sup> DKO mice still yield clear respiratory and fertility phenotypes (data not shown). Distinct roles of miR-34/449 miRNAs could reflect differential expression rather than target specificities.

In a previous study, miR-449 inhibition alone caused defective ciliogenesis by derepressing Notch1 and Dll1 (ref. 10). Yet normal MCC specification in miR-34/449-deficient MCCs suggest that Notch pathway components, with well-characterized roles in regulating MCC specification<sup>42,43</sup>, may not act as key miR-34/449 targets in motile ciliogenesis. Here we provide molecular and functional evidence demonstrating *cp110* as a major miR-34/449 target. Cp110 levels have to be tightly regulated spatially and temporally during ciliogenesis; and proper Cp110 removal from mother centrioles is essential for ciliation<sup>19</sup>. Previous studies mostly focused on ubiquitin-mediated proteasomal degradation of Cp110 (refs 44, 45). Our study reveals the important post-transcriptional regulation of Cp110 by miR-34/449. Although MCC-enriched miR-34/449 miRNAs repress *cp110* to facilitate ciliation, other miRNAs (for example, miR-129) repress *cp110* expression in other cell types to regulate ciliogenesis of primary or motile mono-cilia<sup>36</sup>. Thus, ciliogenesis



**Figure 5 | miR-34/449 miRNAs are required for ciliogenesis by repressing *cp110*.** **a**, *Cp110* mRNA is derepressed in TKO tracheal epithelia ( $n = 4$ ). Paired  $t$ -test: NS,  $P > 0.05$ ,  $*P < 0.05$ ; error bars represent s.e.m. **b**, *Cp110* protein is elevated in TKO tracheal epithelia ( $n = 3$ ). **c**, **d**, miR-34/449 represses *cp110* to regulate MCC ciliation and basal body maturation and docking. **c**, Co-injection of miR-34/449 and *cp110* MOs rescues MCC ciliation, whereas *cp110*Δ3'UTR overexpression phenocopies miR-34/449 morphants. *cp110* MO alone also induced ciliation defects. Ac-α-tub, cilia; phalloidin-488, actin.

is controlled by downregulation of *cp110* through distinct miRNAs in distinct cell types.

For miRNAs, their small size and imperfect target recognition facilitate regulation of multiple mRNA targets. Here, *cp110* knockdown restored basal body maturation/docking and ciliogenesis in miR-34/449-deficient MCCs; however, additional miR-34/449 targets probably exist to regulate the organization of the apical actin cytoskeleton in MCCs.

Mucociliary epithelia are morphologically and functionally conserved among vertebrates<sup>27</sup>. Our findings demonstrated a conserved mechanism that regulates basal body maturation and docking in MCCs by miR-34/449-dependent fine-tuning of *Cp110* levels. This mechanism could have profound implications for the underlying mechanisms disrupted in patients with PCD-like syndromes.

## METHODS SUMMARY

Detailed information on the following methods can be found in the full Methods: mouse breeding and monitoring, histological analyses, qRT-PCR, *in situ* hybridization (ISH), visualization of ciliary beating and mucociliary transport, immunofluorescence staining, scanning and transmission electron microscopy (SEM and TEM), western blotting and manipulation of *Xenopus* embryos.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 January; accepted 29 April 2014.

- Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).

For quantification see Extended Data Fig. 8b. **d**, Basal body maturation/docking is re-established in miR-34/449 morphants upon *cp110* knockdown.

Centrin4-RFP, basal bodies; phalloidin-488, actin. Insets show apical actin meshwork. Total numbers of embryos/cells analysed: Ctrl MO (2/3), miR-34/449 MOs (5/10), miR-34/449 MOs plus *cp110* MO (4/7). Embryos were derived from at least two females and independent fertilizations per *Xenopus* experiment. **e**, Proposed model of regulatory role of miR-34/449 during MCC ciliogenesis through direct repression of *cp110*.

- He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nature Rev. Genet.* **5**, 522–531 (2004).
- Kim, V. N. Small RNAs: classification, biogenesis, and function. *Mol. Cell* **19**, 1–15 (2005).
- Du, T. & Zamore, P. D. microPrimer: the biogenesis and function of microRNA. *Development* **132**, 4645–4652 (2005).
- Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
- Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).
- Miska, E. A. et al. Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. *PLoS Genet.* **3**, e215 (2007).
- Park, C. Y. et al. A resource for the conditional ablation of microRNAs in the mouse. *Cell Rep.* **1**, 385–391 (2012).
- Marson, A. et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521–533 (2008).
- Marcet, B. et al. Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway. *Nature Cell Biol.* **13**, 693–699 (2011).
- He, L. et al. A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130–1134 (2007).
- Chang, T. C. et al. Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol. Cell* **26**, 745–752 (2007).
- Raver-Shapira, N. et al. Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Mol. Cell* **26**, 731–743 (2007).
- He, L., He, X., Lowe, S. W. & Hannon, G. J. microRNAs join the p53 network—another piece in the tumour-suppression puzzle. *Nature Rev. Cancer* **7**, 819–822 (2007).
- Hermeking, H. p53 enters the microRNA world. *Cancer Cell* **12**, 414–418 (2007).
- Satir, P. & Christensen, S. T. Overview of structure and function of mammalian cilia. *Annu. Rev. Physiol.* **69**, 377–400 (2007).
- Fliegauf, M., Benzing, T. & Omran, H. When cilia go bad: cilia defects and ciliopathies. *Nature Rev. Mol. Cell Biol.* **8**, 880–893 (2007).
- Spektor, A., Tsang, W. Y., Khoo, D. & Dynlacht, B. D. Cep97 and CP110 suppress a cilia assembly program. *Cell* **130**, 678–690 (2007).

19. Tsang, W. Y. & Dynlacht, B. D. CP110 and its network of partners coordinately regulate cilia assembly. *Cilia* **2**, 9 (2013).
20. Choi, Y. J. *et al.* miR-34 miRNAs provide a barrier for somatic cell reprogramming. *Nature Cell Biol.* **13**, 1353–1360 (2011).
21. Lizé, M., Klimke, A. & Döbelstein, M. MicroRNA-449 in cell fate determination. *Cell Cycle* **10**, 2874–2882 (2011).
22. Loges, N. T. *et al.* *DNAI2* mutations cause primary ciliary dyskinesia with defects in the outer dynein arm. *Am. J. Hum. Genet.* **83**, 547–558 (2008).
23. Castleman, V. H. *et al.* Mutations in radial spoke head protein genes *RSPH9* and *RSPH4A* cause primary ciliary dyskinesia with central-microtubular-pair abnormalities. *Am. J. Hum. Genet.* **84**, 197–209 (2009).
24. Stubbs, J. L., Oishi, I., Izpisua Belmonte, J. C., Kintner, C. & Izpisu, J. C. The forkhead protein Foxj1 specifies node-like cilia in *Xenopus* and zebrafish embryos. *Nature Genet.* **40**, 1454–1460 (2008).
25. Marshall, W. F. Basal bodies: platforms for building cilia. *Curr. Top. Dev. Biol.* **85**, 1–22 (2008).
26. Avasthi, P. & Marshall, W. F. Stages of ciliogenesis and regulation of ciliary length. *Differentiation* **83**, S30–S42 (2012).
27. Werner, M. E. & Mitchell, B. J. Understanding ciliated epithelia: the power of *Xenopus*. *Genesis* **50**, 176–185 (2012).
28. Gomperts, B. N., Gong-Cooper, X. & Hackett, B. P. Foxj1 regulates basal body anchoring to the cytoskeleton of ciliated pulmonary epithelial cells. *J. Cell Sci.* **117**, 1329–1337 (2004).
29. Klos Dehning, D. A. *et al.* Deuterosome-mediated centriole biogenesis. *Dev. Cell* **27**, 103–112 (2013).
30. Martinez-Anton, A. *et al.* Changes in microRNA and mRNA expression with differentiation of human bronchial epithelial cells. *Am. J. Respir. Cell Mol. Biol.* **49**, 384–395 (2013).
31. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
32. Miranda, K. C. *et al.* A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell* **126**, 1203–1217 (2006).
33. Hoh, R. A., Stowe, T. R., Turk, E. & Stearns, T. Transcriptional program of ciliated epithelial cells reveals new cilium and centrosome components and links to human disease. *PLoS ONE* **7**, e52166 (2012).
34. Tanos, B. E., Yang, H. & Soni, R. Centriole distal appendages promote membrane docking, leading to cilia initiation. *Genes Dev.* **27**, 163–168 (2013).
35. Lai, Y. *et al.* Inflammation-mediated upregulation of centrosomal protein 110, a negative modulator of ciliogenesis, in patients with chronic rhinosinusitis. *J. Allergy Clin. Immunol.* **128**, 1207–1215.e1 (2011).
36. Cao, J. *et al.* miR-129–3p controls cilia assembly by regulating CP110 and actin dynamics. *Nature Cell Biol.* **14**, 697–706 (2012).
37. Park, T. J., Mitchell, B. J., Abitua, P. B., Kintner, C. & Wallingford, J. B. Dishevelled controls apical docking and planar polarization of basal bodies in ciliated epithelial cells. *Nature Genet.* **40**, 871–879 (2008).
38. Delaval, B., Covassin, L., Lawson, N. D. & Doxsey, S. Centrin depletion causes cyst formation and other ciliopathy-related phenotypes in zebrafish. *Cell Cycle* **10**, 3964–3972 (2011).
39. Okada, N. *et al.* A positive feedback between p53 and miR-34 miRNAs mediates tumor suppression. *Genes Dev.* (2014).
40. Kott, E. *et al.* Loss-of-function mutations in *RSPH1* cause primary ciliary dyskinesia with central-complex and radial-spoke defects. *Am. J. Hum. Genet.* **93**, 561–570 (2013).
41. Kunitomo, K. *et al.* Coordinated ciliary beating requires Odf2-mediated polarization of basal bodies via basal feet. *Cell* **148**, 189–200 (2012).
42. Deblandre, G. A., Wettstein, D. A., Koyano-Nakagawa, N. & Kintner, C. A two-step mechanism generates the spacing pattern of the ciliated cells in the skin of *Xenopus* embryos. *Development* **126**, 4715–4728 (1999).
43. Tsao, P. N. *et al.* Notch signaling controls the balance of ciliated and secretory cell fates in developing airways. *Development* **136**, 2297–2307 (2009).
44. D'Angiolella, V. *et al.* SCF<sup>Cyclin F</sup> controls centrosome homeostasis and mitotic fidelity through CP110 degradation. *Nature* **466**, 138–142 (2010).
45. Li, J. *et al.* USP33 regulates centrosome biogenesis via deubiquitination of the centriolar protein CP110. *Nature* **495**, 255–259 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank M. J. Bennett, M. Butler, B. Dynlacht, W. Finkbeiner, P. Kysar, B. Lee, T. Machen, B. Mitchell and J. Wallingford for constructs, technical assistance, stimulating discussions and helpful input. We also thank P. Margolis for careful reading of our manuscript. L.H. acknowledges an R01 and an R21 grant from NCI (R01 CA139067, 1R21CA175560-01), a CIRM new faculty award (RN2-00923-1), a TRDRP research grant (21RT-0133), and a research scholar award from American Cancer Society (ACS, 123339-RSG-12-265-01-RMC). R.S. acknowledges the support of Siebel postdoctoral fellowship and CIRM postdoctoral fellowship. P.W. was funded by the Deutsche Forschungsgemeinschaft (DFG, Wa 3365/1-1), and frog work in the Harland laboratory was funded by NIH grant GM42341. M.L. would like to thank M. Döbelstein for support and discussions, and was financed by a Dorothea Schloerzer Fellowship.

**Author Contributions** R.S. identified and characterized PCD-like motile cilia defects in *mir-34/449* TKO mice, defined *mir-34/449* expression in ciliated epithelia, identified and validated *cp110* as a key *mir-34/449* target in mice. P.W. contributed all *Xenopus* data, in particular providing functional data validating *cp110* as a key *mir-34/449* target. N.S. performed all immunofluorescence experiments in mice, and contributed to target validation experiments. R.S. and P.W. both made significant contribution to experimental planning and result interpretation. M.L., A.K. and M.K. generated and characterized *mir-449* KO mice, defined *mir-449* expression patterns in mouse embryos and contributed to the revision of the manuscript. J.S.L. and G.D. contributed to histology analyses and qRT-PCR analyses. P.L. contributed to the high-speed imaging experiments. Y.W. characterized *mir-34/449* expression in human respiratory epithelia. R.H. contributed to the interpretation of data and to manuscript preparation. L.H. generated *mir-34a* and *mir-34b/34c* KO mice, interpreted the data and coordinated with different groups to complete this study. R.S., P.W., N.S. and L.H. were the major contributors to the preparation of this manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.H. ([lhe@berkeley.edu](mailto:lhe@berkeley.edu)) and R.S. ([rui.song@berkeley.edu](mailto:rui.song@berkeley.edu)).



## METHODS

**Mouse breeding, genotyping and monitoring.** *mir-34a*<sup>-/-</sup>; *mir-34b/34c*<sup>+/-</sup>; *mir-449*<sup>-/-</sup> intercross mating and *mir-34a*<sup>-/-</sup>; *mir-34b/34c*<sup>-/-</sup>; *mir-449*<sup>+/-</sup> intercross mating were both established to generate *mir-34/449* triple knockout (TKO) mice. TKO mice were generated on a mixed genetic background containing C57BL/6, 129 and CD1, and were subsequently backcrossed to C57BL/6 for at least four generations. All mice were housed in a non-barrier animal facility at UC-Berkeley. The following primers were used for genotyping, with parenthetical values indicating the size of the diagnostic PCR product: *mir-34a*-Common-R, ACTGCTGTACCC TGCTGCTT, with *mir-34a*-WT-F, GTACCCGACATGCAAACTT (wild-type band, 400 bp), or *mir-34a*-KO-F, GCAGGACCACTGGATCATTT (KO band, 263 bp); *mir-34b/34c*-Common-R, GAGATTTTCGTGGCGCTTTA, with *mir-34b/34c*-WT-F, GCGTCTGTGAATCGTCATT (wild-type band, 264 bp), or *mir-34b/34c*-KO-F, GCAGGCGCATTAACCTTCGTAT (KO band, 155 bp); *mir-449*-Common-R, AC ATCCCCAAGATATCCCA, with *mir-449*-WT-F, GTATCCACGCCACCACA (wild-type band, 724 bp), or *mir-449*-KO-F, GAGTTTCTGGGCTTGCC, (KO band, 406 bp). Litters were monitored daily for the first 60 days for survival, and their body weight was measured every other day for the first 30 days. The sound wave analyses were performed using audios that recorded the TKO phenotypes (Supplementary Video 1), and respiratory sound was analysed using Audacity.

**Histological analyses.** Tissues were dissected and fixed overnight in 10% neutral buffered formalin, pH 7.4 (NBF) (Fisher Scientific, #SF100-4), processed by standard procedures, and embedded in paraffin blocks. All the blocks were sectioned at 10 µm, and slides were stained by hematoxylin and eosin (H & E). In this analysis, lungs were inflated via trachea with 10% NBF before fixation; and sinuses were processed by post-fixation decalcification for 5 to 10 days in 10% EDTA, pH 7.0.

**qRT-PCR.** Total RNA was isolated by TRIzol (Invitrogen, #15596) from tracheal epithelium per the manufacturer's protocol, and treated with DNase I (Invitrogen, #18068) to remove DNA contamination. For quantitation of mRNAs, TRIzol prepared RNA was reversely transcribed into cDNA using SuperScript III reverse transcriptase (Invitrogen, #18080) with random primers. SYBR Green-based qPCR was subsequently performed on a 7900HT real-time PCR system (Applied Biosystems) using cDNA as template. The *Gapdh*-encoding transcript was used as an endogenous control in each qPCR. The following qPCR primers were used in this study, with parenthetical values indicating the size of the diagnostic PCR product: *Cp110*-F: TCTCCACTGCTTACCATTGA, and *Cp110*-R: GTAAATGGTTTCTGTTGC CC (195 bp); *Foxj1*-F: CTCCTATGCCACTCTCATCT, and *Foxj1*-R: GGATGGA ATTCTGCCAGGTG (137 bp); *Gapdh*-F: AACTTTGGCATTGTGGAAGG, and *Gapdh*-R: CACATTGGGGTAGGAACAC (222 bp). Four independent groups of mice were collected for qRT-PCR quantitation of *Cp110* and *Foxj1*. Each group contained a TKO mouse, an age-matched wild-type mouse, and a littermate-controlled DKO mouse (*mir-34a*<sup>-/-</sup>; *mir-34b/34c*<sup>-/-</sup> or *mir-34a*<sup>-/-</sup>; *mir-449*<sup>-/-</sup>). For *Xenopus laevis* qRT-PCR, cDNA was generated from total RNA extracts of skin explants using iScript reverse transcription supermix (BioRad, #170-8840); and the following qPCR primers were used: *Foxj1*-F: CCAGTGATAGCAAAAGAGGT, and *Foxj1*-R: GC CATGTTCTCCTAATGGAT; *Cp110*-F: AGCCAGAATCCAAGTAAAGG, and *Cp110*-R: CTTGCTTCTTTTCAGAGTC; *Efla*-F: CCCTGCTGGAGCTCTT GAC, and *Efla*-R: GGACACCACTCTCCACAGA; *ODC*-F: GGGCTGGATC TATCTGTAGA, and *ODC*-R: TGCCAGTGTGGTCTTGACAT. Reactions were performed on a BioRad CFX96 Real-Time System C1000 Touch.

For miRNA quantitation, TRIzol prepared total RNA was poly (A)-tailed by Poly (A) Polymerase (Epicentre, #PAP5104H). Poly (A)-tailed small RNA was reversely transcribed into small RNA cDNA with SuperScript III reverse transcriptase (Invitrogen, #18080) using miRNA RT primer (5'-CGAATCTAGAGCTCGAGGCCA GGCGACATGGCTGGCTAGTTAAGCTTGGTACCGAGCTCGGATCCACTA GTCTTTTTTTTTTTTTTTTTTTTTTTVN-3'). V is A, G or C; N is A, G, C or T. TaqMan-based qPCR was subsequently performed on a 7900HT fast real-time PCR system (Applied Biosystems). The U6 snRNA was used as the endogenous control for miRNA real-time qPCR analyses. Universal TaqMan probe, CT CGGATCCACTAGTC; Universal reverse primer, CGAATCTAGAGCTCGAG GCAG. The following forward primers, specific for each small RNA, were used in our studies: *miR-34a*, TGGCAGTGTCTTAGCTGGTTGT; *miR-34b*, AGGCAG TGTAATTAGCTGATTGT; *miR-34c*, AGGCAGTGTAGTTAGCTGATTGC; *miR-449a*, TGGCAGTGTATTGTTAGCTGGT; *miR-449b*, AGGCAGTGTGTTAG CTGGC; *miR-449c*, AGGCAGTGCATTGCTAGCTGG; and U6 snRNA, CGCA AATTCGTGAAGCGTTCC. For *Xenopus* miRNA quantitation the following primers were used: U6 snRNA: ATGTGAAGCGTTCCATATGA; *miR-34a*: TGGCA GTGCTTAGCTGGTTGTT; *miR-34b*: CAGGCAGTGTAGTTAGCTGATTG; *miR449c*: TGCACCTTGCTAGCTGGCTGT. Statistical evaluation was performed using a paired *t*-test.

**In situ hybridization (ISH).** Standard histology protocols were used to prepare P25 lung and trachea section for miRNA ISH using diethylpyrocarbonate (DEPC) treated water for all procedures. After deparaffinization and rehydration, slides were

fixed with 4% paraformaldehyde (PFA), treated with proteinase K, and fixed again with 4% PFA. Slides were incubated first with pre-hybridization solution (3 to 4 h at 60 °C), and then with hybridization solution mixed with digoxigenin (DIG)-labelled LNA probes against each miR-34/449 miRNA (16 h at 60 °C). Post hybridization, slides were washed for 10 min at 60 °C in a graded series of SSC solutions (2×, 1.5×, 0.2×), then incubated with alkaline phosphatase (AP)-conjugated anti-DIG antibody in blocking solution. After washing with PBS and alkaline phosphatase (AP) buffer, the slides were incubated with NBT/BCIP in AP buffer to visualize blue ISH signals. Nuclear fast red (Sigma, #N3020) was used for nuclear counter-staining. Slides then were dehydrated and mounted with Permount (Fisher Scientific, #SP15-100). Solutions for ISH (BioChain #K2191020) as described above. DIG-labelled miR-34a (ACAACCAGCTAAGACACTGCCA), miR-34c (GCAATCAGCTAACTAC ACTGCCT), and miR-449c (CCAGTGAACATGCACTGCCT) probes were purchased from Exiqon (#38487-01, #38542-01, and #39641-01 respectively). For *Xenopus laevis* ISH, embryos were fixed in MEMFA at the indicated stages, and standard protocols were used for ISH and bleaching of embryos<sup>46</sup>. *Foxj1* antisense probe<sup>24</sup> was synthesized using SP6 polymerase (Promega, #P1085). Whole mount ISH was performed on groups of 25 control and manipulated specimens per time point and batch, which were derived from two different mothers.

**Visualization of ciliary beating and mucociliary transport.** Trachea from adult mouse was cut into 2 mm × 2 mm pieces under dissection microscope. Trachea pieces then were transferred into the chamber on a glass slide, which was made by placing a 0.5-mm sticky spacer (Bio-Rad, #SLF-1201) on the slide surface. The chamber was filled with 100 µl M199 Hank's balanced salts medium (Invitrogen, #12350-039) mixed with 1 µl red fluorescent 0.5-µm microspheres (Invitrogen, #F-8812); and a coverglass was placed on the sticky spacer to seal the chamber. Live images of the tracheal epithelium were recorded with a high-speed GX-1 Memrecam camera (NAC Image Technology) attached to an Olympus IX71 microscope. DIC channel was used to record multiciliary beating, and the red fluorescent channel was used to record mucociliary transport. Videos were recorded at 250 frames per second (FPS) for 8 s, and are played at 250 FPS in Supplementary Videos 2 and 3. ImageJ was used to process and analyse raw images (Supplementary Videos 2 and 3).

**Immunofluorescence staining and confocal imaging.** For immunofluorescence staining on cryosections, whole tracheas of adult mice were fixed overnight in ice-cold acetone, and then processed through a graded series of sucrose solutions (from 5% to 20%, Fisher Scientific, #S5-500). Tracheas were embedded in (1:1) 20% sucrose and O.C.T. compound (Tissue-Tek, #4583) and sectioned with MICROM HM 550 (Fisher Scientific) at -21 °C at a thickness of 6 µm. Slides were washed in PBS (3×, 15 min), blocked (1 h at room temperature) in PBSTB (0.1% Triton X-100, 1% bovine serum albumin in PBS), and incubated (overnight at 4 °C) with primary antibodies (1:400, anti-*Foxj1*, Sigma HPA005714; 1:1000 anti-acetylated-α-tubulin, Sigma T6793). Slides then were washed three times in PBST (0.1% Triton X-100 in PBS) and incubated (1 h at room temperature) with secondary antibody (1:1000, Cy3-goat-anti-mouse, Molecular Probes A10521 or 1:500 Alexa Fluor 488 goat anti-rabbit, Molecular Probes A11034). Slides were mounted with VECTASHIELD mounting medium with DAPI (Vector Laboratories, #H-1200). Images were taken with a Zeiss LSM 710 AxioObserver inverted 34-channel confocal microscope and analysed with Zeiss Zen software.

For whole trachea staining, tracheas were cut longitudinally into two pieces, which were fixed either in 80% methanol (EMD, #MX0485P-4) with 20% DMSO (Fisher Scientific, #BP231-100) (overnight at -20 °C), or in 4% paraformaldehyde (PFA) (overnight at 4 °C), respectively. Fixed tissues were washed and blocked as described above, and then incubated (overnight at 4 °C) with primary antibodies (1:500 anti-γ-tubulin, Sigma T5192 for the methanol fixed tissue; 1:400 anti-*Foxj1*, Sigma HPA005714 and 1:1000 anti-acetylated-α-tubulin for the PFA fixed tissue). Tissue then was washed, incubated with secondary antibody, counterstained with DAPI, and imaged as described above.

For *Xenopus laevis* MCC staining, immunofluorescence was performed on whole-mount embryos and skin explants fixed at embryonic stages 30–33 (unless specified otherwise) in 4% paraformaldehyde for 1–2 h at room temperature<sup>47</sup> or in Dent's fixative (80% methanol, 20% dimethylsulfoxide) for 48 h at -20 °C. Embryos were processed according to standard procedures<sup>46</sup>. Morphological analysis of cell types was performed as previously described<sup>48</sup>. Primary antibodies were as follows: mouse monoclonal anti-acetylated-α-tubulin (1:700; Sigma T6793); rabbit polyclonal anti-γ-tubulin (1:500; Sigma T5192). Secondary antibodies (1:250) were as follows: AlexaFluor 488-labelled goat anti-mouse antibody (Molecular Probes A11001), AlexaFluor 555-labelled goat anti-mouse antibody (Molecular Probes A21422), AlexaFluor 555-labelled goat anti-rabbit antibody (Molecular Probes A21428) and AlexaFluor 405-labelled goat anti-mouse antibody (Molecular Probes A31553). Actin staining was performed by incubation (30–60 min at room temperature) with AlexaFluor 488-labelled Phalloidin (1:40; Molecular Probes A12379). Z-stack analysis and processing were performed using ImageJ and Zeiss ZEN software. All confocal imaging was performed using a Zeiss LSM700.

**Scanning electron microscopy (SEM).** Adult trachea tissue was fixed using Karnovsky's fixative in 0.1 M sodium phosphate buffer (Sorenson's), washed with Sorenson's sodium phosphate buffer and post-fixed using 1% OsO<sub>4</sub> in Sorenson's for 1 h. Tissue was dehydrated by passing through a graded series of ethanol solutions, then critical point dried using a Tousimis 931 super critical point dryer. The tissue was mounted on aluminium stubs and sputter coated with gold using a PELCO SC-7 coater. The samples were viewed on an FEI XL30 TMP SEM and digital images were collected. SEM was performed in the electron microscopy facility of the University of California at Davis.

**Transmission electron microscopy (TEM).** Adult trachea tissue was fixed, washed, and post-fixed as for SEM. After rinsing in double-distilled water (DDW), the tissue was incubated (30 min at room temperature) in 0.1% tannic acid, rinsed again in DDW, and incubated (1 h) in 1% uranyl acetate in DDW. Tissue was dehydrated by passing through a graded series of acetone solutions, then infiltrated and embedded in an epoxy resin mixture. Survey thick sections were cut, and ultra-thin sections of the selected areas were generated with a diamond knife (Diatome). The thin sections were picked up on copper grids and stained with uranyl acetate and lead citrate before viewing on a Philips CM120 Biotwin. Micrographs were taken with a Gatan MegaScan Model 794/20 digital camera. TEM analyses with longitudinal and transverse sections were performed in the electron microscopy facility of the University of California at Davis. ImageJ was used to measure the distance between basal bodies and apical surface in TEM pictures, and Oriana was used to analyse the directionality of ciliary axonemes. The following criteria were used to determine multiciliated cells (MCCs) and their apical surface in longitudinal TEM. Cells containing basal bodies in tracheal epithelium were defined as MCCs. For ciliated MCCs, the surface with intact view from ciliary axoneme to basal body was determined as the apical surface; and for non-ciliated MCCs, the surface with microvilli was determined as the apical surface.

**Air-liquid interface (ALI) culture of primary tracheal epithelia.** Primary tracheal epithelial cells were cultured as described previously<sup>49</sup>. In short, tracheas from three adult *mir-34/449* mutant mice of the same genotype (~P60) were dissected and cut longitudinally in ice-cold Ham's F-12 medium (Life Technologies, #11765-054) with penicillin/streptomycin (Life Technologies, #15140-163). To isolate epithelial cells, tracheas were incubated in 1.5 mg ml<sup>-1</sup> pronase (Roche Diagnostics, #101659 21001) in Ham's F-12 media overnight at 4 °C. We stopped the tracheal pronase digestion by adding fetal bovine serum (FBS, Omega Scientific, #FB-01) to a final concentration of 10%. The tracheas were washed twice in ice-cold fresh Ham's F-12 media with 10% FBS. Tracheal epithelial cells were then pelleted by pooling the pronase digestion and washes for centrifugation at 400g for 10 min at 4 °C. The pelleted cells were treated by DNase I (Sigma, #DN25), resuspended in 1 ml FBS, then plated in 9 ml pre-warmed basic medium (BM), which contained 1:1 DMEM: Ham's F-12 (Life Technologies, #11330-032), penicillin/streptomycin, 1.5 mM glutamine (Life Technologies, #25030-149), 0.03% sodium bicarbonate (Life Technologies, #25080-094), and 0.1% fungizone (Life Technologies, #15290-018). Fibroblasts were depleted by a 4 h incubation at 37 °C, after which the tracheal epithelial cells were plated on trans-wells (Corning, #3470) in 24-well plates in proliferation medium, which contained BM with 5% FBS, 25 ng ml<sup>-1</sup> epidermal growth factor (EGF) (BD Biosciences, #354001), 10 µg ml<sup>-1</sup> insulin (Sigma, #I1882), 5 µg ml<sup>-1</sup> apo-transferrin (Sigma, #T1147), 0.1 µg ml<sup>-1</sup> cholera toxin extract (Sigma, #C8052), 30 µg ml<sup>-1</sup> bovine pituitary extract (BPE) (Hammond Cell Tech, #1078-NZ), and 50 nM retinoic acid (RA, Sigma, #R2625-500MG). The proliferation medium was changed every other day for 7 days, until we switched the culture to air-liquid interface culture by removing the media in the apical chamber, and replacing the media in the basal chamber with differentiation medium (BM with 2% Nu-Serum (BD, #355100) and 50 nM RA). The differentiation medium was changed every other day for 23 days, when cells were ready for immunofluorescence staining.

**Luciferase assay.** A fragment of the *Cp110* mRNA 3' UTR containing two predicted miR-34/449 sites was cloned into the FseI site immediately downstream of the stop codon in the pGL3-Control firefly luciferase vector (Promega, #E1741). We amplified a fragment of *Cp110* 3' UTR using PCR with *Cp110*-3' UTR-F, AAGGC CGGCCGAAGACAGCACTCACTGGGA, and *Cp110*-3' UTR-R, GTGGCCGG CCTTCTCTGAGATCCGGATTGC. NIH/3T3 cells were cultured in 10% bovine serum in DMEM (Invitrogen, #11995-073) in a 12-well plate at a density of 1 × 10<sup>5</sup> cells per well. We co-transfected each well of NIH/3T3 cells with 10 ng of pGL3 constructs, 100 ng of pRL-TK *Renilla* vector (Promega, #E2241), and 15 nM miR-34b miRNA mimics (Integrated DNA Technologies, 5'-AGGCAGUGUAAU AGCUGAUUGU-3' and 5'-AAUCACUACUCCACUGUUAUC-3') or siGFP (Integrated DNA Technologies) using TransIT-TKO Transfection Reagent (Mirus

Bio, # MIR 2150). At 24 h after transfection, firefly and *Renilla* luciferase activities were measured using the Dual-Luciferase reporter assay system (Promega, #E1910). The luciferase activity was normalized as the ratio of firefly/*Renilla* luciferase activities.

**Western blotting.** Protein was collected from mouse tracheal epithelium by incubating resected and cleaned trachea in radioimmunoprecipitation assay (RIPA) buffer with complete mini protease inhibitor cocktail tablet (Roche, #11836153001) for 30 min on ice. Western blot followed the standard protocols. mouse anti-β-actin (Sigma, #A5441, ~46 kDa) was used as loading control at 1:40,000 dilution. Rabbit anti-Cp110 (Thermo Scientific, # PA5-34380, ~120 kDa) was used at 1:1,000 dilution in 5% non-fat dry milk in TBS-T (20 mM Tris pH 7.6, 150 mM NaCl and 0.1% Tween-20). Horseradish peroxidase (HRP) conjugated secondary antibodies (Santa Cruz Biotechnology, #sc-2004 and #sc-2005) were used at 1:5,000 dilution. Blots were analysed using ImageJ. Band intensities were normalized against corresponding actin and compared to wild-type controls for ratio calculation.

**Manipulation of *Xenopus laevis* embryos and skin explants.** *X. laevis* eggs were collected and *in vitro* fertilized, then cultured and microinjected by standard procedures<sup>46</sup>. Embryos were injected with morpholino nucleotides (MOs, Gene Tools) at the two- to four-cell stage using a PicoSpritzer setup in 1/3× modified frog Ringer's solution (MR) with 2.5% Ficoll PM 400 (GE Healthcare 17-0300-50), and then were transferred into 1/3× MR containing gentamycin<sup>46</sup> after injection. Drop size was calibrated to about 7–8 nl per injection. Rhodamine-B dextran (0.5–1.0 mg ml<sup>-1</sup>; Invitrogen D1841) was co-injected and used as lineage tracer. MO (Gene Tools) doses were administered as follows: a total dose of 30 ng miR-34/449 MOs; 10 ng each for miR-34a MO (5'-CAACAACCAGCTAAGACACTGCCAA-3'), miR-34b MO (5'-ACAATCAGCTAACTACACTGCCTGA-3'), and miR-449a MO (5'-AACCAGCTAACATTACACTGCCTT-3'), or 30 ng of a miR-control MO (5'-TGCACGTTTCAATACAGACCGT-3'), or 17 ng of *cp110* MO (5'-ACTCTT CATATGGCTCCATGGTCCC-3'). An mRNA encoding Centrin-4-RFP or Centrin-GFP<sup>37</sup> and Sas6-GFP<sup>29</sup> was prepared using the Ambion message machine kit using SP6 (AM1340) and diluted to 50–100 ng µl<sup>-1</sup> for injection into the embryos (0.8–1.6 ng total per embryo). *Xenopus tropicalis* *cp110* cDNA (PureYield Midiprep; Promega, #A2495) was derived from a clone matching BC167469 obtained from Thermo Scientific (#MXT1765-202715711). *Cp110*Δ3'UTR was generated from the same clone, which was digested with SphI (New England Biolabs, #R0182S) and re-ligated to remove most of the 3' UTR. Analysis of morphant tadpole brains for signs of hydrocephalus was performed as previously described<sup>50</sup>.

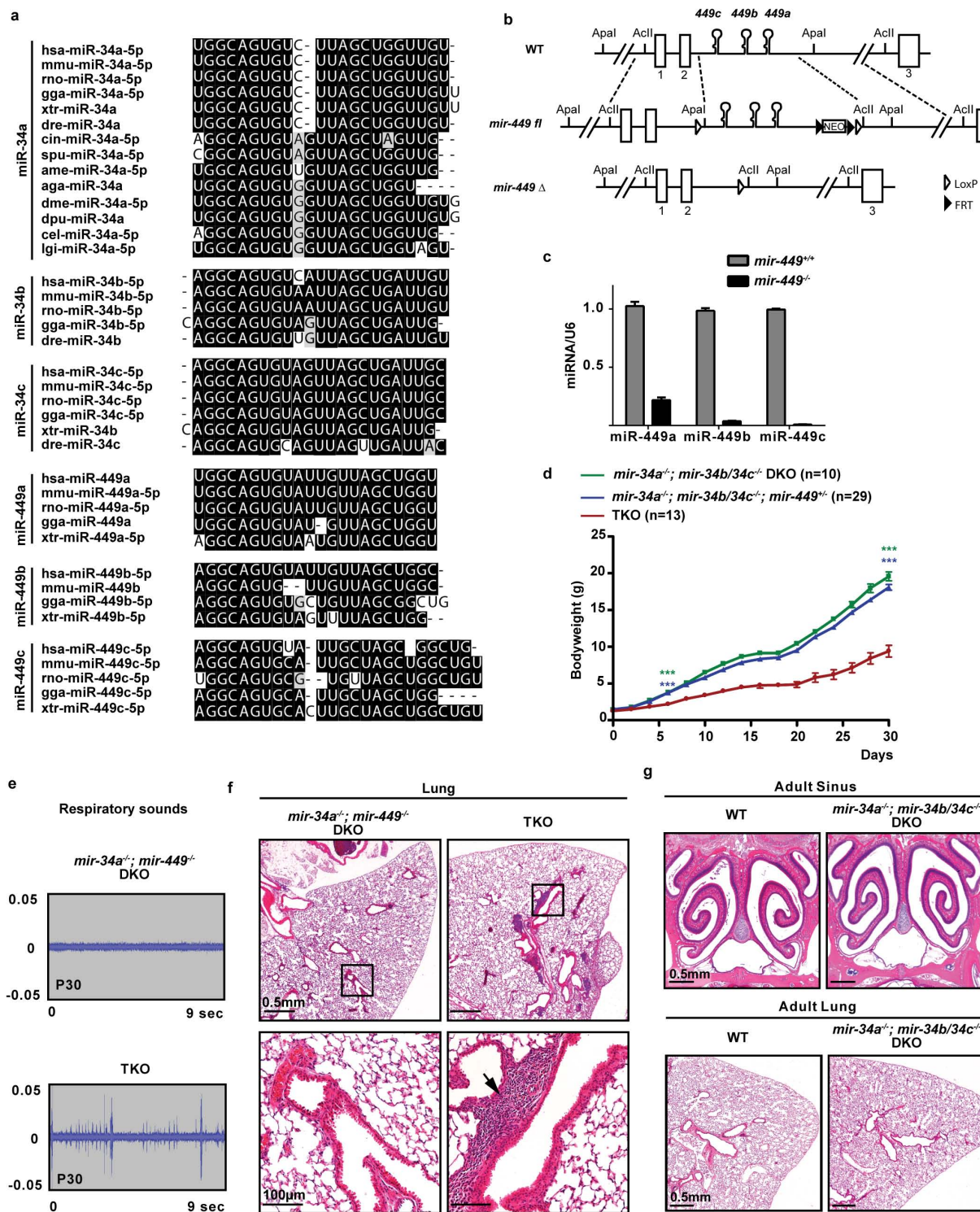
*Xenopus* skin explants were generated from animal caps<sup>46</sup>, dissected in 1× modified Barth's saline from stage 9 embryos, which were either un-injected (time course experiment) or injected with Ctrl MO or miR-34/449 MOs (for quantification of *cp110* and *foxj1* expression). Explants were cultured in 0.5× modified Barth's saline until un-manipulated control embryos reached appropriate stage. In the time course experiment, stage 10, 26 and 32 explants represented ciliation state of MCCs for quantitation of *cp110*, *foxj1* and miR-34/449 miRNAs. In addition, *cp110* and *foxj1* expression levels in Ctrl MO and miR-34/449 MOs injected embryos (injected 4× into the animal hemisphere at the 4-cell stage) were assessed at onset of ciliation (stage 26 explants) to examine the effect of miR-34/449 on *foxj1* and *cp110* levels. Statistical evaluation of experimental data was performed using chi-squared tests (<http://www.physics.csbsju.edu/stats/contingency.html>) or Wilcoxon sum of ranks (Mann-Whitney) tests ([http://www.fon.hum.uva.nl/Service/Statistics/Wilcoxon\\_Test.html](http://www.fon.hum.uva.nl/Service/Statistics/Wilcoxon_Test.html)).

**Sample size and analysis.** Sample sizes for all experiments were chosen based on previous experiences. No randomization or blinding was applied for all our studies.

**Ethics statement on animal experiments.** This work was done with approval of University of California, Berkeley's Animal Care and Use Committee. University of California, Berkeley's assurance number is A3084-01, and is on file at the National Institutes of Health Office of Laboratory Animal Welfare.

46. Sive, H. L., Grainger, R. M. & Harland, R. M. *Early Development of Xenopus laevis* (Cold Spring Harbor Laboratory Press, 2000).
47. Walentek, P., Beyer, T., Thumberger, T., Schweickert, A. & Blum, M. ATP4a is required for Wnt-dependent Foxj1 expression and leftward flow in *Xenopus* left-right development. *Cell Rep.* **1**, 516–527 (2012).
48. Walentek, P. *et al.* A novel serotonin-secreting cell type regulates ciliary motility in the mucociliary epidermis of *Xenopus* tadpoles. *Development* **141**, 1526–1533 (2014).
49. Vlado, E. K. & Brody, S. L. Analysis of ciliogenesis in primary culture mouse tracheal epithelial cells. *Methods Enzymol.* **525**, 285–309 (Elsevier, 2013).
50. Hagenlocher, C., Walentek, P., Müller, C., Thumberger, T. & Feistel, K. Ciliogenesis and cerebrospinal fluid flow in the developing *Xenopus* brain are regulated by *foxj1*. *Cilia* **2**, 12 (2013).





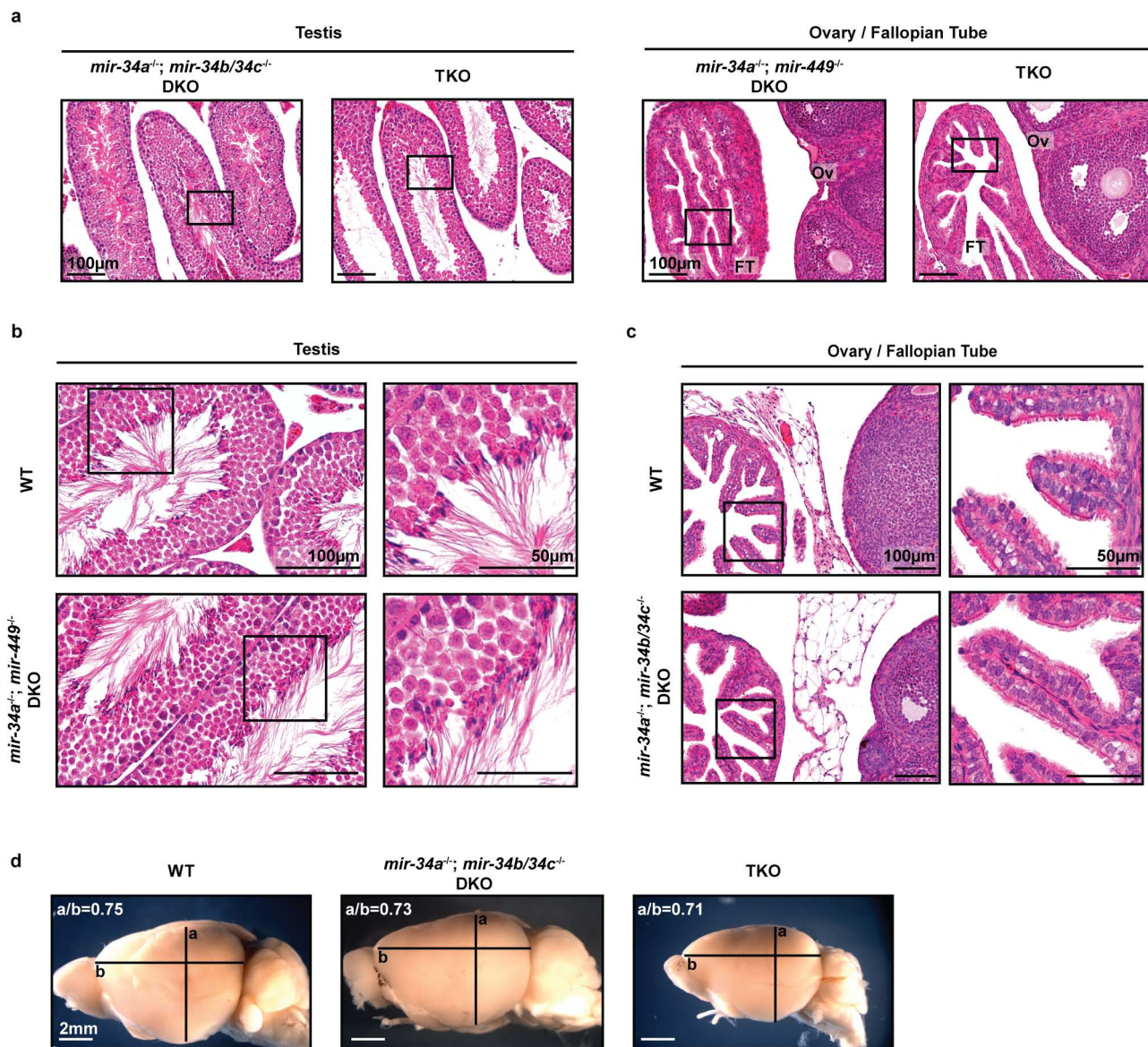
Extended Data Figure 1 | The generation and phenotypic characterization

**of *mir-34/449* TKO mice.** **a**, *mir-34/449* miRNAs are evolutionarily conserved with extensive sequence homology across many species. *mir-34a* has a more ancient evolutionary history compared to the rest of *mir-34/449* miRNAs. *mir-34a* is conserved in Deuterostome, Ecdysozoa and Lophotrochozoa, yet the rest of *mir-34/449* miRNAs have only vertebrate homologues. **b**, Diagrams of the targeted deletion strategy to generate *mir-449* knockout mice. Since all *mir-449* miRNAs are within intron 2 of their host gene, *cdc20b*, we deleted *mir-449* with a minimally predicted impact on *cdc20b*. **c**, *mir-449* expression is absent in *mir-449*<sup>-/-</sup> knockout animals, as demonstrated by real-time PCR analyses in lung tissues from littermate-controlled wild-type and *mir-449*<sup>-/-</sup> mice at postnatal day 35 (*n* = 3). *mir-449a* real-time PCR primers exhibit a modest cross-reaction with *mir-34*

miRNAs. **d**, *mir-34/449* TKO mice have a significant postnatal attenuation in body weight. Littermate-controlled *mir-34a*<sup>-/-</sup>; *mir-34b/34c*<sup>-/-</sup>, *mir-34a*<sup>-/-</sup>; *mir-34b/34c*<sup>-/-</sup>; *mir-449*<sup>+/-</sup> and TKO mice were monitored for their body weight every other day for 30 days after birth. Paired *t*-test, \*\*\**P* < 0.001.

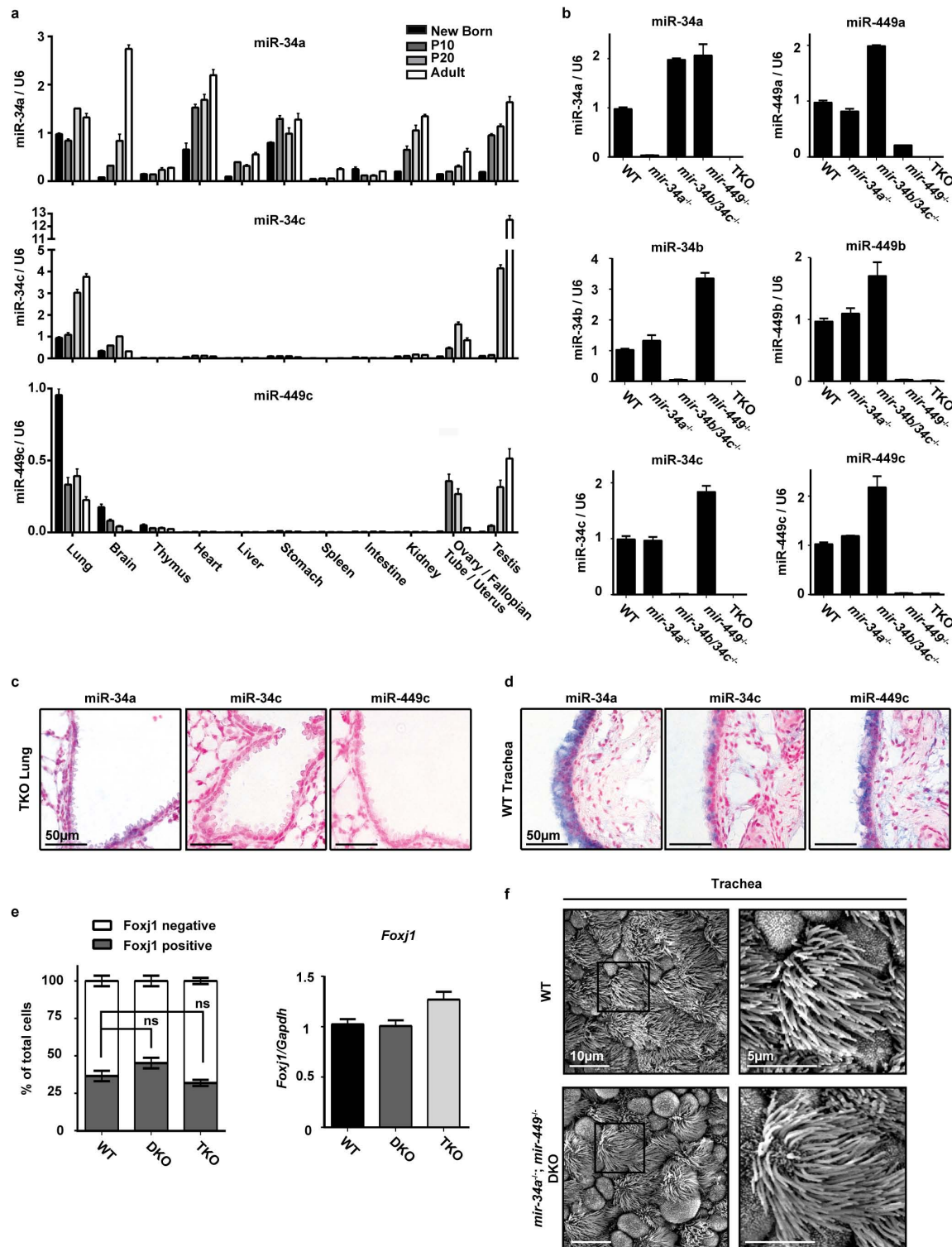
**e**, Surviving *mir-34/449* TKO mice exhibit coughing/sneezing-like phenotype. The respiratory noise of littermate-controlled *mir-34a*<sup>-/-</sup>; *mir-449*<sup>-/-</sup> DKO and TKO mice was shown by sound wave analysis at postnatal day 30 (*n* = 14). **f**, Pulmonary inflammation occurs in a subset of *mir-34/449* TKO mice. A representative H&E analysis of lung tissues from an adult TKO mouse indicates an increased infiltration by inflammatory cells. A total of 3 out of 15 TKO mice examined exhibit lung infection. **g**, *mir-34a*<sup>-/-</sup>; *mir-34b/34c*<sup>-/-</sup> DKO mice resemble wild-type mice, exhibiting no obvious respiratory defects in paranasal sinus or lung (*n* = 3). All error bars represent s.e.m.





**Extended Data Figure 2 | Phenotypic characterization of reproductive organs and brain in *mir-34/449* TKO mice.** **a**, Adult male and female *mir-34/449* TKO mice are infertile. Male (left) and female (right) reproductive organs from littermate-controlled DKO and TKO mice were subjected to H&E staining ( $n = 3$ ). Boxes indicate areas depicted in Fig. 1e. **b**, Adult *mir-34a<sup>-/-</sup>; mir-449<sup>-/-</sup>* DKO male mice exhibit no defects in spermatogenesis. **c**, Adult

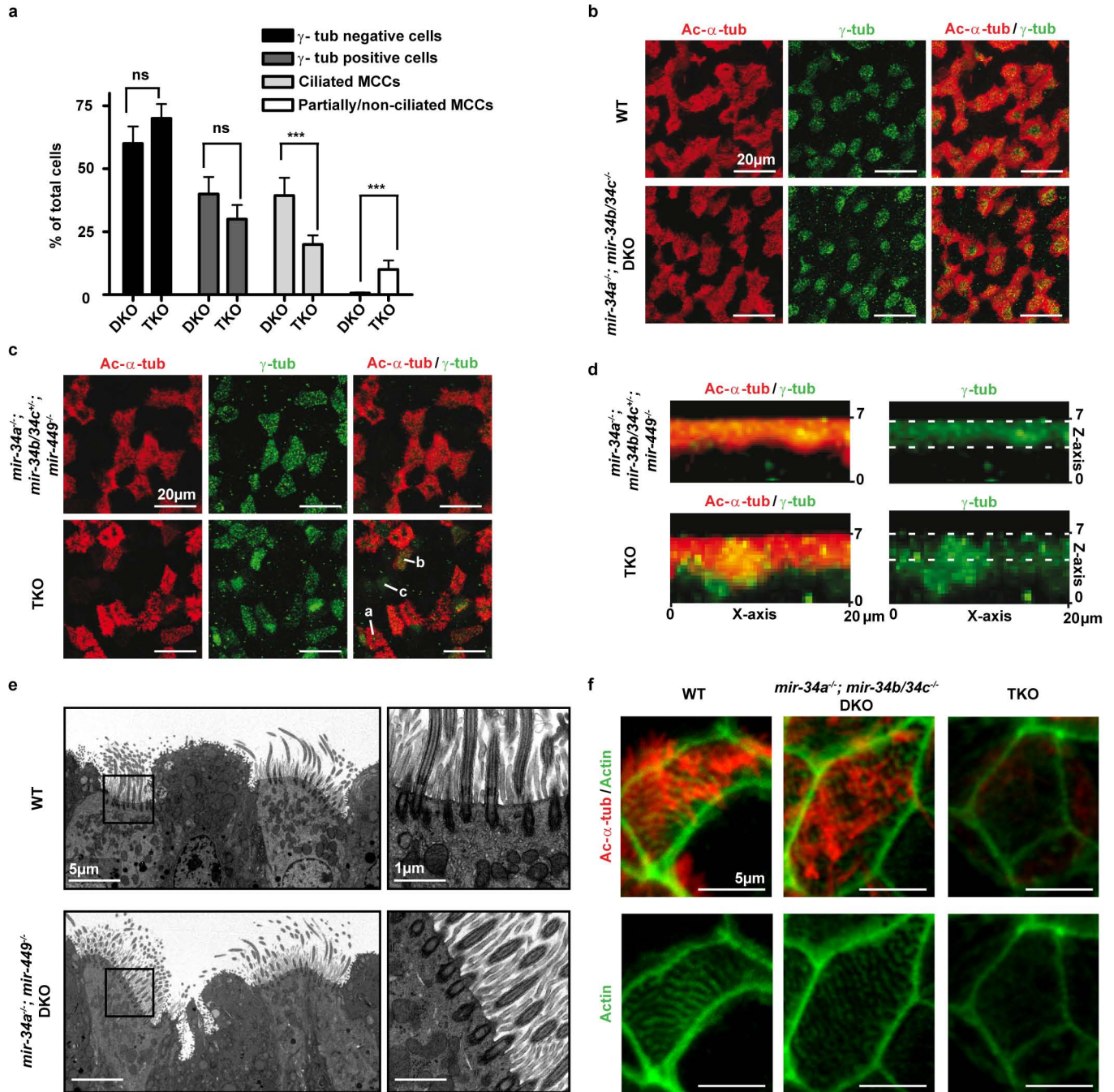
*mir-34a<sup>-/-</sup>; mir-34b/34c<sup>-/-</sup>* DKO female mice display no defects in reproductive organs. **d**, The adult *mir-34/449* TKO brains do not exhibit hydrocephalus, yet they are smaller in size than wild-type and DKO controls. The a/b markings on the images indicate the coronal to horizontal ratios.  $n = 3$  for **b**, **c** and **d**.



**Extended Data Figure 3 | miR-34/449 miRNAs are enriched in airway MCCs.** **a**, Most miR-34/449 miRNAs are enriched in tissues with motile cilia. Using real-time PCR, the expression of miR-34a, miR-34c and miR-449c was measured in multiple tissues from newborn, P10, P20 and adult wild-type mice. Both miR-34c and miR-449c are exclusively expressed in tissues with motile cilia, whereas miR-34a exhibits a broader expression pattern ( $n = 3$ ). **b**, The real-time PCR assay for each miR-34/449 miRNA specifically detects the corresponding miRNA. The specificity of each miRNA real-time PCR assay was validated using testis RNA from wild-type (WT), *mir-34a*<sup>-/-</sup>, *mir-34b/34c*<sup>-/-</sup>, *mir-449*<sup>-/-</sup>, and TKO mice at postnatal day 35. The miR-449a assay shows a slight cross reaction with homologous miRNAs ( $n = 3$ ). **c**, *In situ* hybridization of each miR-34/449 miRNA exhibits specific detection. No

measurable miR-34/449 *in situ* signal is detected in TKO lung sections at postnatal day 25 ( $n = 2$ ). **d**, miR-34/449 miRNAs are enriched in tracheal MCCs. *In situ* hybridization analyses demonstrate that miR-34c and miR-449c are specifically expressed in the tracheal MCCs, whereas miR-34a is expressed in both tracheal MCCs and the surrounding cell types ( $n = 2$ ). **e**, *mir-34/449* TKO mice do not exhibit significant alterations in Foxj1 expression. Quantification of Foxj1 positive cells (left,  $n = 3$ ) and Foxj1 mRNA (right,  $n = 4$ ) was performed for well-controlled wild-type, DKO and TKO tracheas, using immunofluorescence and real-time PCR, respectively. Paired *t*-test; ns,  $P > 0.05$ . **f**, *mir-34a*<sup>-/-</sup>; *mir-449*<sup>-/-</sup> DKO tracheal epithelia are morphologically indistinguishable from wild-type controls in scanning electron microscopy (SEM) analyses ( $n = 3$ ). All error bars represent s.e.m.

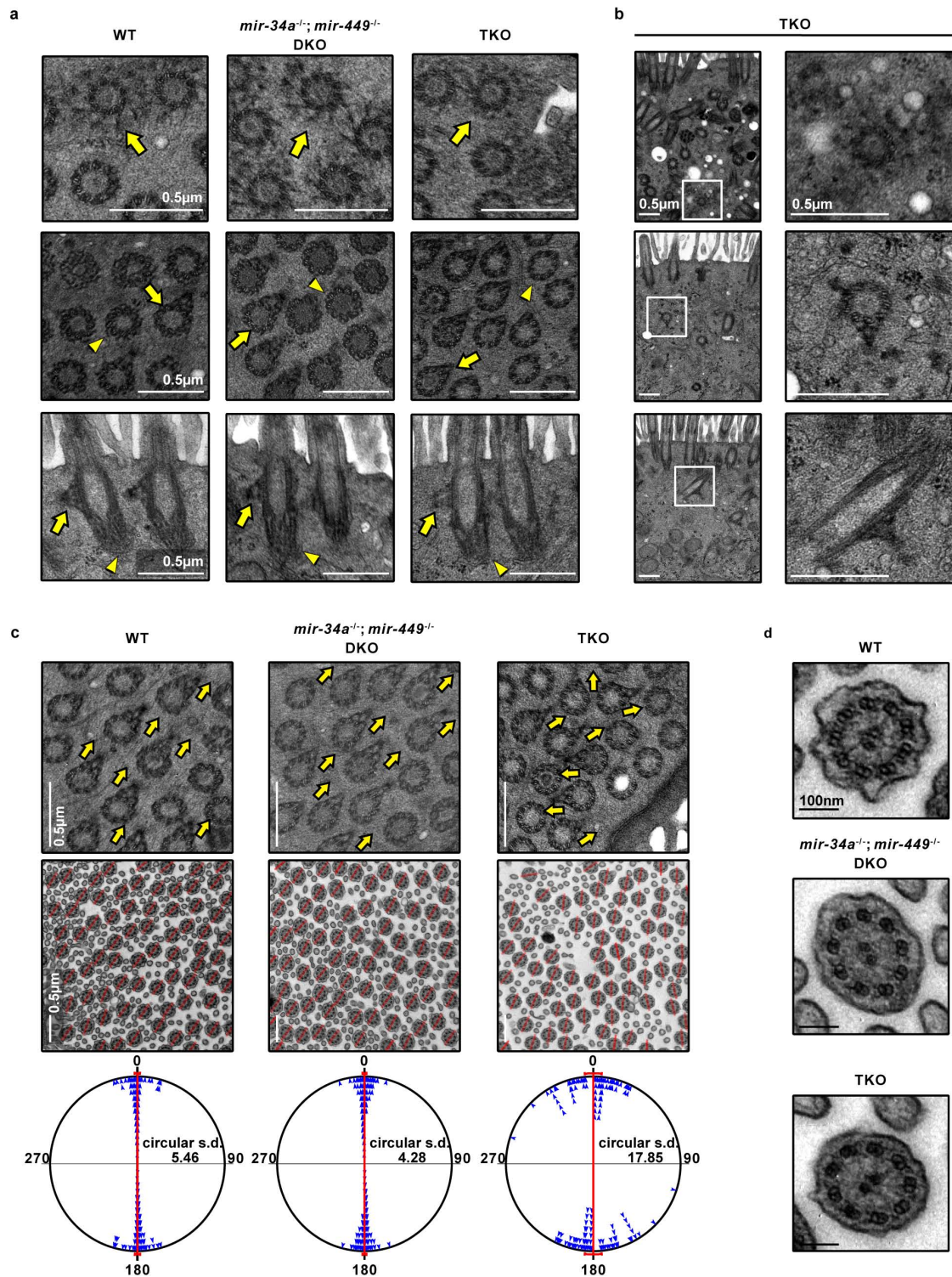




**Extended Data Figure 4 | miR-34/449 deficiency causes defective ciliation and basal body docking in mouse airway MCCs.** **a**, *mir-34/449* TKO trachea exhibit reduced MCC ciliation. Quantification of fully ciliated MCCs ( $\gamma$ -tub and Ac- $\alpha$ -tub double-positive) and partially/non-ciliated MCCs, Ac- $\alpha$ -tub weak(negative) was performed in littermate controlled DKO and TKO mouse tracheas, using data from all three experiments in Fig. 3a. The number of cells with MCC identity ( $\gamma$ -tub positive) is unaffected in TKO tracheas, yet one-third of the TKO MCCs display aberrant Ac- $\alpha$ -tub staining, indicating ciliation defects. Paired *t*-test; ns,  $P > 0.05$ , \*\*\* $P < 0.001$ . **b**, The *mir-34a<sup>-/-</sup>; mir-34b/34c<sup>-/-</sup>* DKO mice exhibit normal ciliogenesis in tracheal MCCs. Whole tracheas from age-matched adult wild-type and *mir-34a<sup>-/-</sup>; mir-34b/34c<sup>-/-</sup>* DKO mice were analysed by immunofluorescence staining for Ac- $\alpha$ -tub (cilia) and  $\gamma$ -tubulin (basal bodies) ( $n = 3$ ). **c**, *mir-34/449* TKO primary tracheal epithelial cells exhibit ciliation defects in air liquid interface (ALI) culture. ALI culture of MCCs were derived from tracheas of littermate-controlled *mir-34a<sup>-/-</sup>; mir-34b/34c<sup>-/-</sup>; mir-449<sup>-/-</sup>* and TKO mice, and

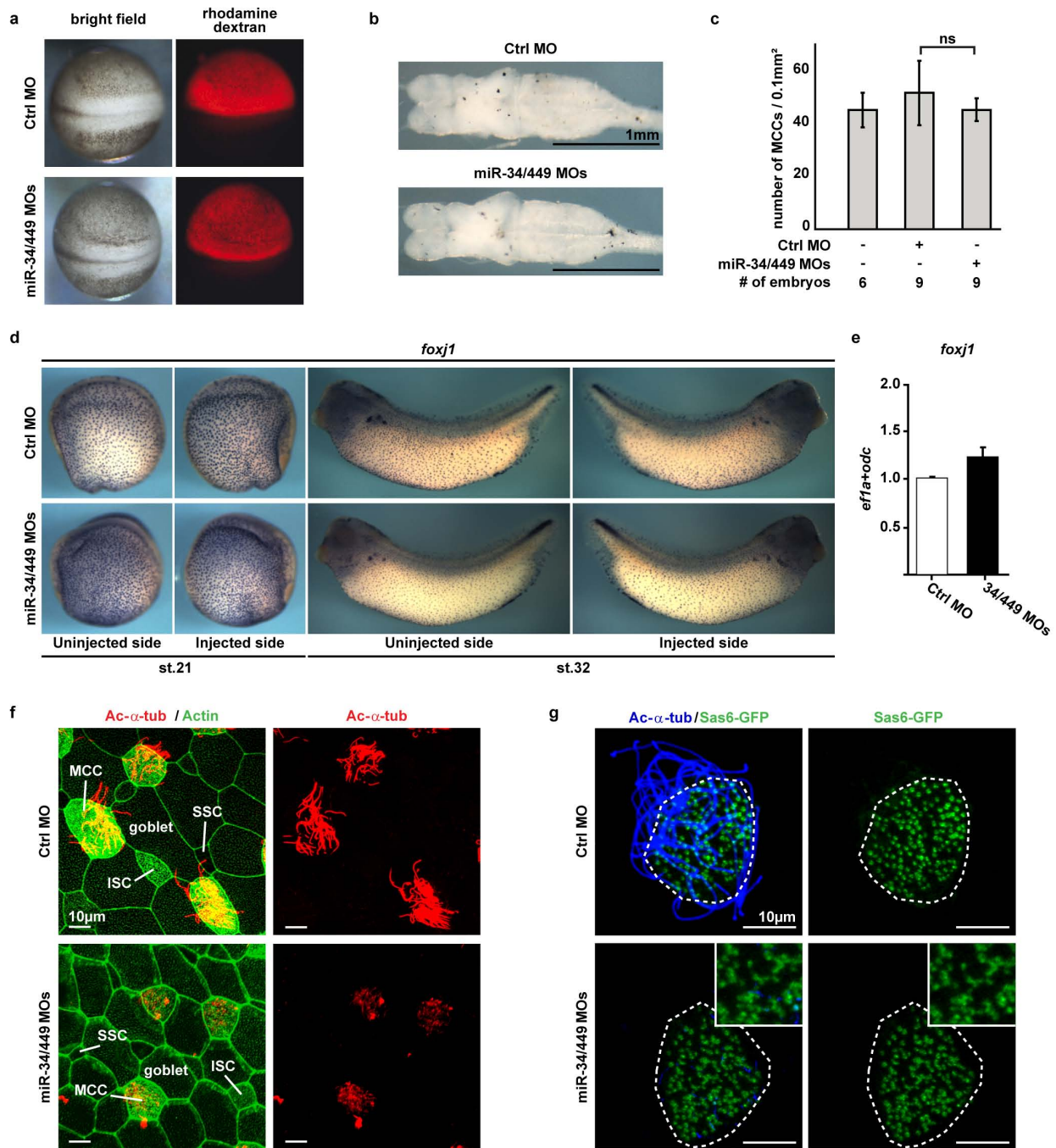
subjected to immunofluorescence staining for Ac- $\alpha$ -tub (cilia) and  $\gamma$ -tub (basal bodies). In TKO and control ALI culture, comparable levels of  $\gamma$ -tub positive cells are observed, however a large portion of TKO  $\gamma$ -tub positive cells displayed a partial or complete loss of Ac- $\alpha$ -tub staining. The letters on the image indicate fully (a), partially (b) or non-ciliated (c) MCCs ( $n = 2$ ). **d**, Basal bodies fail to dock to the apical membrane of *mir-34/449* TKO MCCs in ALI culture. Lateral projections of confocal micrographs described in (c) show impaired apical localization of  $\gamma$ -tub staining in TKO MCCs from ALI cultures, suggesting a defective basal body docking to the apical membrane. **e**, *mir-34a<sup>-/-</sup>; mir-449<sup>-/-</sup>* DKO trachea exhibit no defects in basal body docking using transmission electron microscopy (TEM) analyses ( $n = 3$ ). **f**, TKO tracheal MCCs exhibit a defective subapical actin network. Whole tracheas from adult wild-type, *mir-34a<sup>-/-</sup>; mir-34b/34c<sup>-/-</sup>* DKO and TKO mice were analysed by immunofluorescence staining for Ac- $\alpha$ -tub (cilia) and phalloidin-488 (actin) ( $n = 2$ ). All error bars represent s.e.m.





**Extended Data Figure 5 | Major basal body structural components are intact in *mir-34/449* TKO MCCs revealed by transmission electron microscopy.** **a**, **b**, Apically docked (**a**) and undocked (**b**) basal bodies in *mir-34/449* TKO MCCs have intact structural components. Basal body transition fibres (top), basal feet (middle) and striated rootlets (bottom) have comparable morphology among WT, DKO and TKO MCCs. For the top panel an arrow indicates a representative transverse view of transition fibres. For the middle panel an arrow indicates a representative transverse view of nine microtubule triplets with basal feet and arrowhead indicates a representative transverse view collected from a different height of a basal body, containing nine microtubule triplets without basal feet. For the bottom panel, an arrow

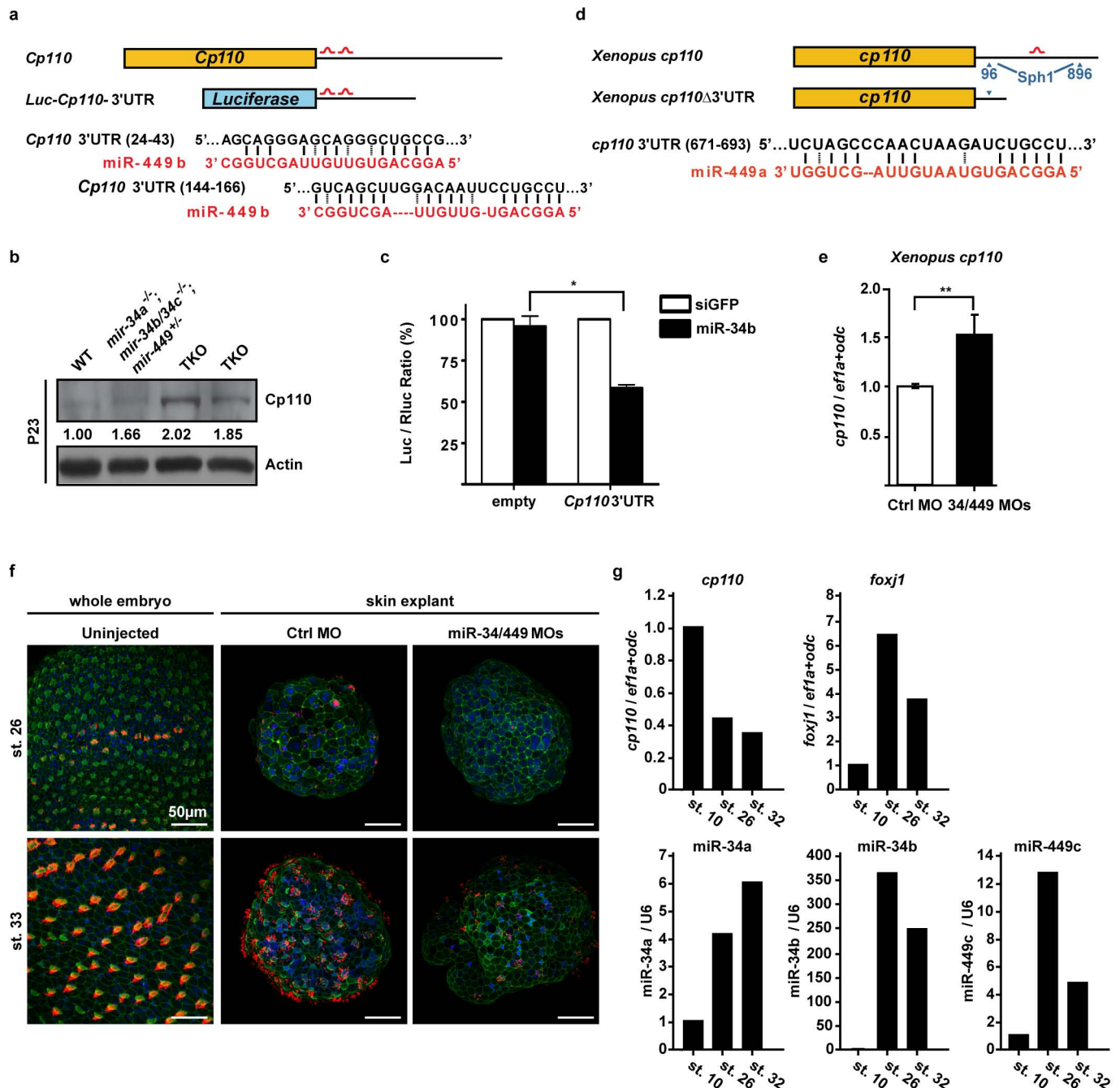
indicates the longitudinal view of basal feet and arrowhead indicates the striated rootlet structure. **c**, Directionality of basal bodies (top) and axonemes (middle) is moderately affected in *mir-34/449* TKO MCCs. In the top panel arrows point to the directions indicated by basal foot. In the middle panel red lines connecting the central pair of axonemes indicate the rotational polarity of each ciliary axoneme. In the bottom panel the angles of the axoneme directionality were statistically analysed as bidirectional circular data. The average angel was set from 0° to 180° axis. *mir-34/449* TKO ciliary axonemes have moderately un-coordinated directionality compared to WT and DKO controls. **d**, *mir-34/449* TKO axonemes exhibit intact structures, including nine outer microtubule doublets, two central microtubule singlets and dynein arms ( $n = 3$ ).



**Extended Data Figure 6 | miR-34/449 deficiency in frog MCCs causes defective ciliogenesis without affecting cell fate specification.** **a**, Injection of Ctrl or miR-34/449 MOs does not affect general embryonic development or neural tube closure. *Xenopus laevis* embryos were injected unilaterally with MOs at the 2–4-cell stage and analysed at neurula stages (18–20). Targeting of the skin ectoderm was confirmed by co injection of fluorescent rhodamine dextran. **b**, Frog miR-34/449 morphants do not exhibit hydrocephalus. Embryos were injected animally with control or miR-34/449 MOs into both dorsal blastomeres at the 4-cell stage to target the neural tube and brain regions. Subsequently, the whole brains were dissected and analysed at stage 45/46. The lack of hydrocephalus in miR-34/449 morphants argues against a role of miR-34/449 in ependymal ciliation. **c**, Quantification of fully ciliated, partially ciliated or non-ciliated MCCs reveals no significant change in total number of MCC-fated cells in miR-34/449 morphants. Error bars represent s.d. Wilcoxon two-sample test, ns,  $P > 0.05$ . **d, e**, *foxj1* expression and specification of MCC fate is unaltered in miR-34/449 deficient embryos. **d**, Embryos were unilaterally injected with Ctrl or miR-34/449 MOs to the right side at the 2–4-cell stage, cultured until stage 21 or 32 and processed for *in situ*

hybridization to monitor *foxj1* expression in the mucociliary epithelium of the skin. No change in *foxj1* expression can be detected. **e**, Real-time PCR analysis in Ctrl or miR-34/449 MOs injected skin explants at stage 26 (onset of ciliation) does not indicate reduced expression levels of *foxj1*. Error bars represent s.e.m. **f**, miR-34/449 deficient frog embryos exhibit normal development of mucociliary cell types. Detailed analysis of the embryonic skin at stage 30–32 reveals the presence (specification and intercalation) of all cell types in miR-34/449 morphants, including large goblet cells, small secretory cells (SSC), Ac- $\alpha$ -tub positive ciliated cells (MCC) and non-tubulin enriched ion secreting cells (ISC). **g**, miR-34/449 morphant MCCs exhibit an uneven distribution of basal bodies. *Sas6-gfp* mRNA was injected at the 2–4-cell stage to visualize basal bodies at stage 30–32. In control embryos Sas6-GFP foci are evenly distributed in fully ciliated MCCs, whereas miR-34/449 morphant MCCs are characterized by an uneven distribution and aggregation of basal bodies, which frequently fail to grow cilia (Ac- $\alpha$ -tub staining). Such phenotype is characteristic for basal body docking defects. Total numbers of embryos/cells analysed were Ctrl MO (4/7) and miR-34/449 MOs (6/10).



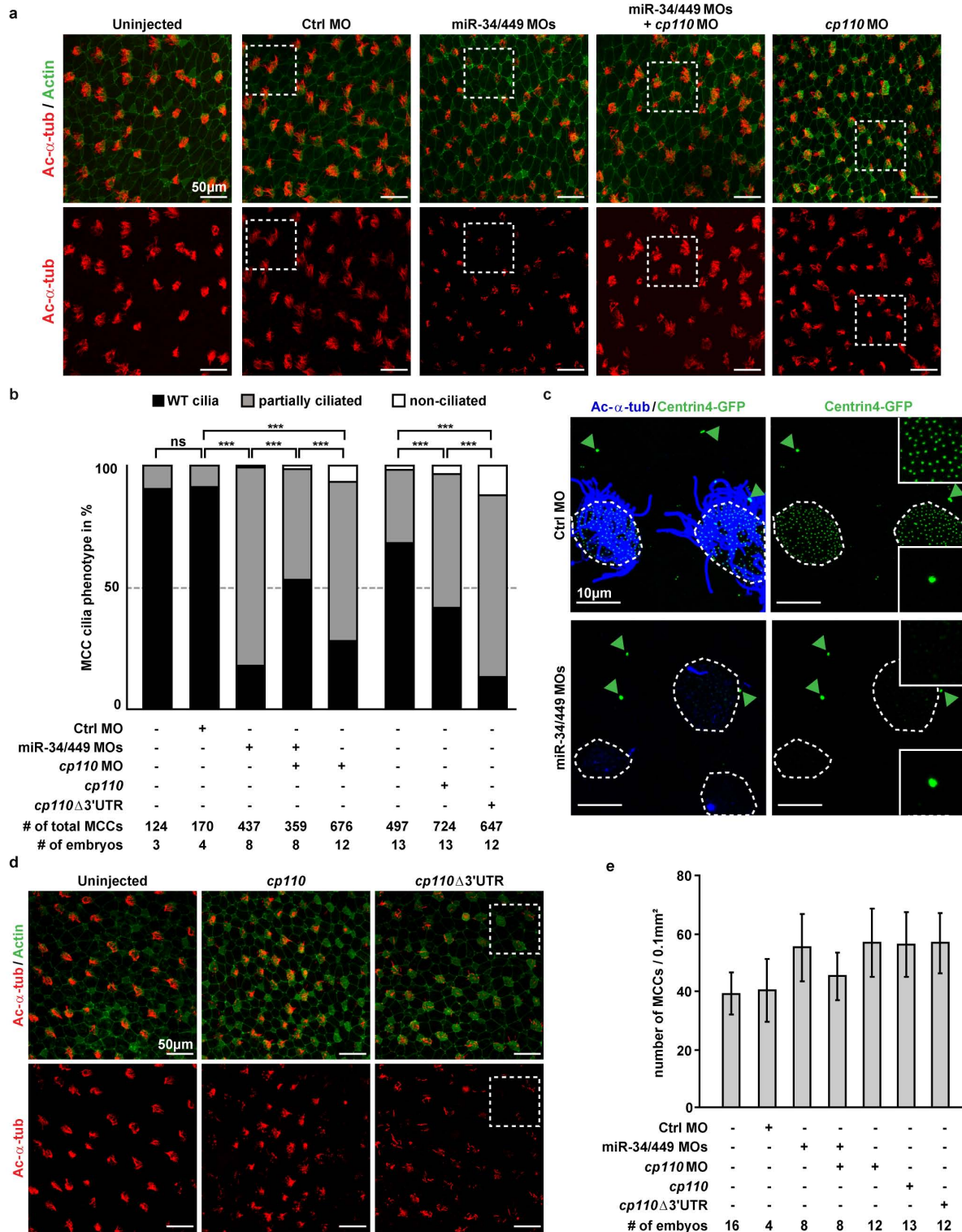


# Extended Data Figure 7 | *cp110* is a direct target of miR-34/449 miRNAs.

**a**, A schematic representation of two predicted miR-34/449 binding sites in the mouse *Cp110* 3' UTR and in the luciferase reporter construct that contains *Cp110* 3' UTR. **b**, *Cp110* protein levels at postnatal day 23 are elevated in *mir-34/449* TKO tracheal epithelia. **c**, The expression of *Luc-Cp110-3'UTR* exhibits miR-34b-dependent repression in NIH/3T3 cells. Error bars represent s.e.m.,  $n = 3$ . Paired  $t$ -test,  $*P < 0.05$ . **d**, A schematic representation of one predicted miR-34/449 binding site in the frog *cp110* 3' UTR. A truncated *cp110* construct, *cp110Δ3'UTR*, was made to generate a *cp110* cDNA without the miR-34/449 target site. **e**, Real-time PCR monitoring *cp110* reveals elevated mRNA expression levels of *cp110* in miR-34/449 morphant frog skin explants as compared to Ctrl MO injected specimens.  $**P < 0.01$ . Paired  $t$ -test. Error bars, s.e.m. **f**, Timeline of MCC ciliation and recapitulation of ciliation defects in skin explants (animal caps). Representative confocal images from staged whole embryos and skin explants injected with either Ctrl MO or miR-34/449

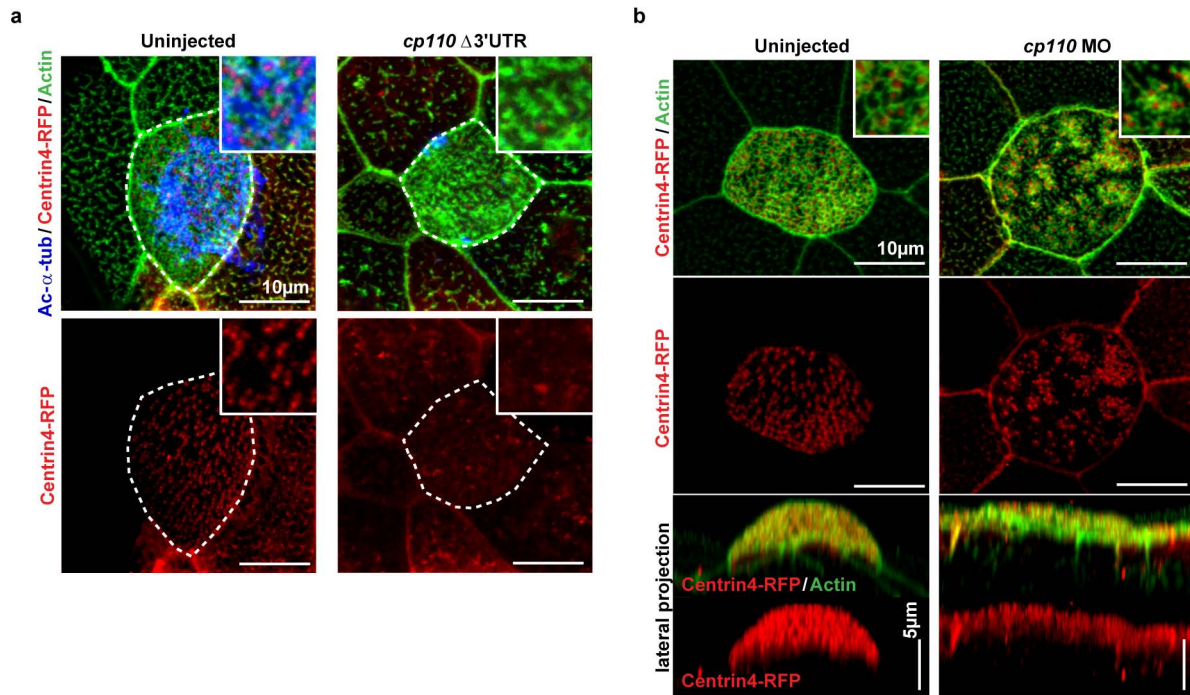
MOs show the onset of ciliation at stage 26 and fully ciliated skin ectoderm at stage 32 in whole embryos and Ctrl MO injected skin explants. miR-34/449 MOs injected skin explants develop MCC ciliogenesis defects comparable to whole embryo treatment. Cilia: Ac- $\alpha$ -Tub (red), Actin: Phalloidin-488 (green) and nuclei: DAPI (blue). **g**, Expression of *cp110*, *foxj1* and miR-34/449 RNAs during time course of ciliation in skin explants. Explants at stage 10 represent unciliated MCC precursors, explants at stage 26 represent MCCs at the onset of ciliation, and stage 32 explants represent fully ciliated ectodermal epithelium. *cp110* mRNA levels decrease over the time course of ciliation, with the strongest decrease between stage 10 and 26, while *foxj1* mRNA levels rapidly increase during this time. miR-34a, -34b and -449c levels strongly increase between stage 10 and stage 26; and only a moderate increase or even decrease can be observed between stage 26 and 32, similar to *foxj1* expression levels. Error bars represent s.e.m.  $n = 2$ , technical replicates on pools of 30 skin explants for each time point.





**Extended Data Figure 8 | miR-34/449 miRNAs promote ciliogenesis by repressing *cp110*.** **a**, Representative examples of confocal images used for quantification of MCC ciliation in **(b)**. Embryos were stained for Ac- $\alpha$ -tub (cilia) and phalloidin-488 (actin). White boxes indicate areas depicted in Fig. 5c. **b**, Quantification of MCC ciliation in **a**, **d** and Fig. 5c.  $\chi^2$ -test, ns  $P > 0.05$ , \*\*\* $P < 0.001$ . **c**, Centrin4-GFP incorporation into basal bodies is affected in miR-34/449 deficient embryos. The *centrin4-gfp* mRNA was injected at the 2–4-cell stage to visualize basal bodies in MCCs at stage 32, and centrosomes in neighbouring epithelial cells. In Ctrl morphant embryos, Centrin4-GFP staining in basal bodies (smaller foci in ciliated cells) and centrosomes (bigger foci in non-ciliated cells, green arrowheads) are equally

strong. In contrast, Centrin4-GFP staining in basal bodies is greatly reduced in miR-34/449 morphants, without alteration of fluorescent intensity in centrosomes of neighbouring cells. Total numbers of embryos/cells analysed were Ctrl MO (6/17), miR-34/449 MOs (7/23). **d**, Representative examples of confocal images from *cp110* overexpression experiments used for quantification of MCC ciliation in **b**. White boxes indicate areas depicted in Fig. 5c. **e**, The number of MCC-fated cells in miR-34/449 or *cp110* morphants, and embryos injected with *cp110* DNA constructs is not reduced. Quantification of total MCC numbers (fully ciliated, partially ciliated or non-ciliated MCCs) is shown for frog embryos injected with various MOs/DNAs (corresponding to **a**, **b**, **d** and Fig. 5c). Error bars represent s.d.



**Extended Data Figure 9 | Gain and loss of *cp110* affects MCC basal bodies, but not the apical actin meshwork.** **a**, *cp110* overexpression phenocopies miR-34/449 knockdown. Centrin4-RFP enrichment is strongly reduced in *cp110* $\Delta 3'$ UTR overexpressing MCCs. It is noteworthy, that whereas ciliation and incorporation of Centrin4 are strongly affected in *cp110* $\Delta 3'$ UTR injected embryos, formation of the apical actin meshwork appears largely unaffected. Together with the lack of *cp110* knockdown to rescue the apical actin meshwork in miR-34/449 morphants, these data indicate an additional effect of miR-34/449 miRNAs on Actin formation/organization, which is *cp110* independent. Cilia: Ac- $\alpha$ -tub, basal bodies: Centrin4-RFP, actin:

phalloidin-488. Embryos/cells analysed: uninjected (4/7), *cp110* $\Delta 3'$ UTR (6/10). **b**, The cellular basis for ciliation defects in *cp110* morphants is probably due to the atypical failure of basal bodies to separate from each other, thus they appear to be aggregated in clusters of *cp110*-deficient MCCs. Nevertheless, Centrin4-RFP incorporation or apical localization of basal bodies is not affected in *cp110* morphants. Basal bodies, Centrin4-RFP; actin: phalloidin-488. Total numbers of embryos/cells analysed: un-injected (2/3), *cp110* MO (5/8). Embryos were derived from at least two females and independent fertilizations per *Xenopus* experiment.

**Extended Data Table 1 | Candidate miR-34/449 targets with potential roles in MCC differentiation and/or ciliation in respiratory epithelia**

Gene Symbol	Gene Name
<i>Ank3</i>	ankyrin 3, epithelial
<i>Atp2b4</i>	ATPase, Ca <sup>++</sup> transporting, plasma membrane 4
<i>Aurka</i>	aurora kinase A
<i>Aurkb</i>	aurora kinase B
<i>Ccdc39</i>	coiled-coil domain containing 39
<i>Ccdc40</i>	coiled-coil domain containing 40
<i>Ccna2</i>	cyclin A2
<i>Ccnb1</i>	cyclin B1
<i>Ccnd2</i>	cyclin D2
<i>Ccne2</i>	cyclin E2
<i>Ccp110</i>	centriolar coiled coil protein 110
<i>Cdc25a</i>	cell division cycle 25A
<i>Cdc6</i>	cell division cycle 6
<i>Cdca7l</i>	cell division cycle associated 7 like
<i>Cdk5rap2</i>	CDK5 regulatory subunit associated protein 2
<i>Cep152</i>	centrosomal protein 152
<i>Cep63</i>	centrosomal protein 63
<i>Cep97</i>	centrosomal protein 97
<i>Chk1</i>	checkpoint kinase 1
<i>Daam1</i>	dishevelled associated activator of morphogenesis 1
<i>Dnah5</i>	dynein, axonemal, heavy chain 5
<i>Dnali1</i>	dynein, axonemal, light intermediate polypeptide 1
<i>Dtl</i>	denticless homolog (Drosophila)
<i>Dzip1</i>	DAZ interacting protein 1
<i>Fat4</i>	FAT tumor suppressor homolog 4 (Drosophila)
<i>Fgfr1</i>	fibroblast growth factor receptor
<i>Foxg1</i>	forkhead box G1
<i>Foxj1</i>	forkhead box J1
<i>Hdac6</i>	histone deacetylase 6
<i>Hook3</i>	hook homolog 3 (Drosophila)
<i>Ift27</i>	intraflagellar transport 27
<i>Itch</i>	itchy, E3 ubiquitin protein ligase
<i>Jag1</i>	jagged 1
<i>Kif24</i>	kinesin family member 24
<i>Lef1</i>	lymphoid enhancer binding factor 1
<i>Mapt</i>	microtubule-associated protein tau
<i>Met</i>	met proto-oncogene
<i>Myb</i>	myeloblastosis oncogene
<i>Myh9</i>	myosin, heavy polypeptide 9, non-muscle
<i>Pacs1</i>	phosphofurin acidic cluster sorting protein 1
<i>Pdgfra</i>	platelet derived growth factor receptor, alpha polypeptide
<i>Pofut1</i>	protein O-fucosyltransferase 1
<i>Rdh11</i>	retinol dehydrogenase 11
<i>Rfx3</i>	regulatory factor X, 3
<i>Rrm2</i>	ribonucleotide reductase M2
<i>Shank3</i>	SH3/ankyrin domain gene 3
<i>Six3</i>	sine oculis-related homeobox 3
<i>Skp2</i>	S-phase kinase-associated protein 2 (p45)
<i>Stat6</i>	signal transducer and activator of transcription 6
<i>Stil</i>	Scl/Tal1 interrupting locus
<i>Stk36</i>	serine/threonine kinase 36
<i>Tmem107</i>	transmembrane protein 107
<i>Tppp</i>	tubulin polymerization promoting protein
<i>Tsc2</i>	tuberous sclerosis 2
<i>Ttc26</i>	tetratricopeptide repeat domain 26
<i>Ttll3</i>	tubulin tyrosine ligase-like family, member 3
<i>Xpnpep3</i>	X-prolyl aminopeptidase (aminopeptidase P) 3, putative



# Crystal structure of the human glucose transporter GLUT1

Dong Deng<sup>1,2,3\*</sup>, Chao Xu<sup>1,2,3\*</sup>, Pengcheng Sun<sup>1,2\*</sup>, Jianping Wu<sup>1,2,3\*</sup>, Chuangye Yan<sup>1,2</sup>, Mingxu Hu<sup>1,2,3</sup> & Nieng Yan<sup>1,2,3</sup>

**The glucose transporter GLUT1 catalyses facilitative diffusion of glucose into erythrocytes and is responsible for glucose supply to the brain and other organs. Dysfunctional mutations may lead to GLUT1 deficiency syndrome, whereas over-expression of GLUT1 is a prognostic indicator for cancer. Despite decades of investigation, the structure of GLUT1 remains unknown. Here we report the crystal structure of human GLUT1 at 3.2 Å resolution. The full-length protein, which has a canonical major facilitator superfamily fold, is captured in an inward-open conformation. This structure allows accurate mapping and potential mechanistic interpretation of disease-associated mutations in GLUT1. Structure-based analysis of these mutations provides an insight into the alternating access mechanism of GLUT1 and other members of the sugar porter subfamily. Structural comparison of the uniporter GLUT1 with its bacterial homologue Xyle, a proton-coupled xylose symporter, allows examination of the transport mechanisms of both passive facilitators and active transporters.**

GLUT1, encoded by *SLC2A1*, mediates the basal-level cellular uptake of glucose into many tissues. In particular, it is responsible for constant uptake of glucose—maintained at approximately 5 mM concentration in blood—into erythrocytes through facilitative diffusion<sup>1–3</sup>. GLUT1 in endothelial cells of the blood–tissue barrier has an essential role for glucose supply to the brain and other organs<sup>4–6</sup>.

Inactivating mutations of GLUT1, resulting in compromised transport activities for glucose, are associated with diseases as a result of lack of energy supply to the brain<sup>7</sup>. GLUT1 deficiency syndrome (also known as De Vivo syndrome) is characterized by a spectrum of symptoms including early-onset seizures, microcephaly and retarded development<sup>8–12</sup>. Cancer cells require enhanced glucose supply in part due to the less efficient energy production through anaerobic glycolysis (the Warburg effect). Elevated expression levels of GLUT1 have been observed in several cancer types, identifying GLUT1 as an important prognostic indicator for tumorigenesis<sup>13–19</sup>. Because of its fundamental physiological and pathophysiological significance, GLUT1 has been a focus for functional study and structural determination<sup>20–22</sup>.

GLUT1 belongs to the sugar porter subfamily of the major facilitator superfamily (MFS), one of the largest and most ubiquitous secondary transporter superfamilies<sup>23,24</sup>. MFS transporters share a conserved core fold that comprises 12 transmembrane segments organized into two discretely folded domains, namely the amino- and carboxy-terminal domains. Within each domain, the six consecutive transmembrane segments are folded into a pair of '3 + 3' inverted repeats<sup>22,25</sup>. Mounting experimental evidence indicates that the three-helix bundle may represent the basic structural and functional unit<sup>25,26</sup>. All MFS transporters are thought to transport substrate using the alternating access mechanism<sup>27</sup>, whereby the substrate-binding site is alternately accessible from either side of the membrane through conformational changes of the transporters.

Structures of bacterial GLUT1 homologues—the D-xylose:H<sup>+</sup> symporter Xyle (refs 28, 29) from *Escherichia coli* and the glucose:H<sup>+</sup> symporter GlcP (ref. 30) from *Staphylococcus epidermidis*—have been reported. Notably, the structures of Xyle bound to D-xylose or D-glucose enabled homology-based modelling of GLUT1 (ref. 28). However, both Xyle

and GlcP are proton-driven symporters whereas GLUT1 is a uniporter that catalyses the translocation of glucose down its concentration gradient across the membrane. Atomic structure of human GLUT1 is essential for understanding its transport and disease mechanisms.

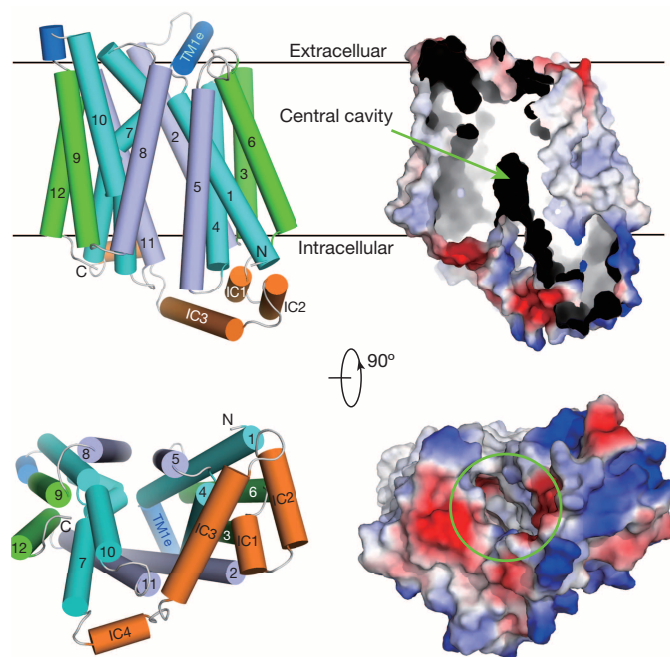
## Structure determination of GLUT1

Details of subcloning, overexpression and purification of GLUT1 can be found in Methods. Full-length human GLUT1 (residues 1–492) with a C-terminal 10× histidine tag was overexpressed in High Five insect cells infected with baculoviruses. To eliminate heterogeneity caused by glycosylation, we introduced a single point mutation (N45T). We tried but were unable to crystallize GLUT1(N45T). We reasoned that GLUT1, a transporter with high activity<sup>31</sup>, may exhibit multiple interchangeable conformations that impede crystallization. To solve this potential problem, we searched the literature for GLUT1 variants that may lock the uniporter in certain conformations. A single missense mutation, E329Q, was previously shown to arrest GLUT1 in an inward-facing conformation<sup>32</sup>. We eventually succeeded in the crystallization of full-length GLUT1 containing the mutations N45T and E329Q. The crystals appeared in the C2 space group and a few of them diffracted X-rays beyond 3.2 Å at synchrotron radiation source. The structure of GLUT1 was determined using molecular replacement with the coordinates of the N and C domains of Xyle as separate searching models. The final atomic model was refined to 3.2 Å resolution (Extended Data Fig. 1 and Extended Data Table 1).

In the GLUT1 structure, residues 9–455 constitute the canonical MFS fold whereas the C-terminal segment was invisible probably due to its inherent flexibility in this conformation (Fig. 1). The N and C domains enclose a cavity that opens to the intracellular side. The structure thereby represents an inward-open conformation, consistent with the predicted state for GLUT1(E329Q)<sup>32</sup>. Notably, the N and C domains are connected by an intracellular helical bundle (ICH) which comprises four short  $\alpha$ -helices (Fig. 1). Intriguingly, the ICH domain was also observed in the structures of the sugar porter members Xyle and GlcP, but not in other MFS transporters, suggesting that ICH may be a unique feature of the sugar porter subfamily<sup>25</sup>.

<sup>1</sup>State Key Laboratory of Bio-membrane and Membrane Biotechnology, Tsinghua University, Beijing 100084, China. <sup>2</sup>Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China. <sup>3</sup>Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing 100084, China.

\*These authors contributed equally to this work.



**Figure 1 | Overall structure of the human glucose transporter GLUT1.** The structure of full-length human GLUT1 containing two point mutations (N45T, E329Q) was determined in an inward-open conformation. The side and cytoplasmic views are shown. The corresponding transmembrane segments in the four 3-helix repeats are coloured the same. The extracellular and intracellular helices are coloured blue and orange, respectively. A slab of cut-open view of the surface electrostatic potential, which was calculated with PyMol<sup>50</sup>, is shown on the right to facilitate visualization of the inward-facing cavity. IC indicates intracellular helix. All structure figures were prepared with PyMol.

### Implications for ligand binding

During refinement of the atomic model, a disc-shaped electron density with a discontinuous tail appeared in the inward-open cavity (Extended Data Fig. 2a). Because GLUT1 was purified and crystallized in the presence of 0.4% (w/v) *n*-nonyl- $\beta$ -D-glucopyranoside ( $\beta$ -NG), the observed electron density may come from a bound  $\beta$ -NG molecule, the sugar moiety of which happens to be the substrate for GLUT1. Indeed, a  $\beta$ -NG molecule fits perfectly into the electron density, with the D-glucopyranoside bound to the C domain of GLUT1 at approximately the centre of the membrane, whereas the aliphatic tail points to the intracellular side along the cavity (Extended Data Fig. 2b, c).

The C domains of GLUT1 and XyleE can be superimposed with a root-mean-squared deviation of 1.244 Å over 138 aligned C $\alpha$  atoms. On domain superimposition, D-glucopyranoside of the bound  $\beta$ -NG overlaps with D-glucose (Extended Data Fig. 2d). Reminiscent of the D-glucose coordination by XyleE (ref. 28), the D-glucopyranoside of  $\beta$ -NG is hydrogen-bonded (H-bonded) to the surrounding polar residues in the C domain (Extended Data Fig. 2e). The similar coordination of the D-glucopyranoside of  $\beta$ -NG by inward-open GLUT1 and D-glucose by outward-facing XyleE supports the notion of a single sugar-binding site that is alternately accessed from either side of the membrane<sup>27,33</sup>. In the outward-facing, partly occluded XyleE, D-glucose is coordinated by residues mainly from the C domain<sup>28</sup>. In the inward-open GLUT1, residues from the N domain are not involved in ligand binding. These observations indicate that the C domain may provide the primary substrate-binding site whereas the relative motion of the N domain results in alternating access.

### Mapping of disease-related mutations

The detailed structural information on GLUT1 provides a molecular basis for mechanistic interpretation of disease-derived inactivating mutations (Extended Data Table 2). These mutations were identified from

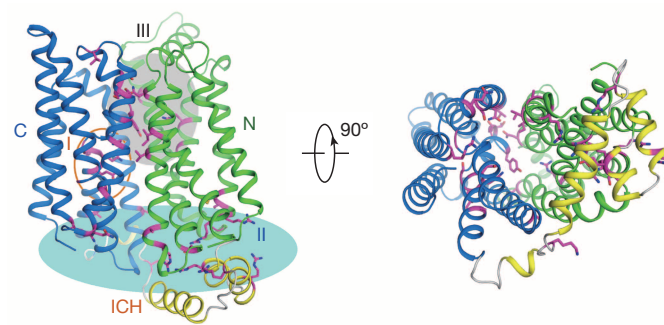
patients with GLUT1 deficiency syndrome, early-onset absence epilepsy, paroxysmal exercise-induced dyskinesia, and familial idiopathic generalized epilepsy<sup>12,34–44</sup>. Most of the mutations target polar or charged amino acids, which probably have a functional, rather than structural, role. It is possible that mutations that disrupt the structural integrity of GLUT1 cause complete loss of function and hence lethality<sup>10</sup>.

The disease-derived mutations largely map to three clusters in GLUT1 (Fig. 2). Cluster I comprises residues that are responsible for substrate binding (Fig. 2 and Extended Data Fig. 2e); cluster II is located at the interface between the transmembrane domain and the ICH domain (Figs 2 and 3); and cluster III comprises residues lining the transport path, which are mostly committed to interactions between the N and C domains on the extracellular side (Figs 2 and 4 and Extended Data Fig. 3a). Mutations of the substrate-binding residues may lead to compromised recognition of D-glucose and markedly reduced transport activity<sup>28,45</sup>. Below, we focus on the analysis of clusters II and III which may facilitate understanding of the transport mechanism of GLUT1 and other sugar porter members.

### Critical role of the ICH domain

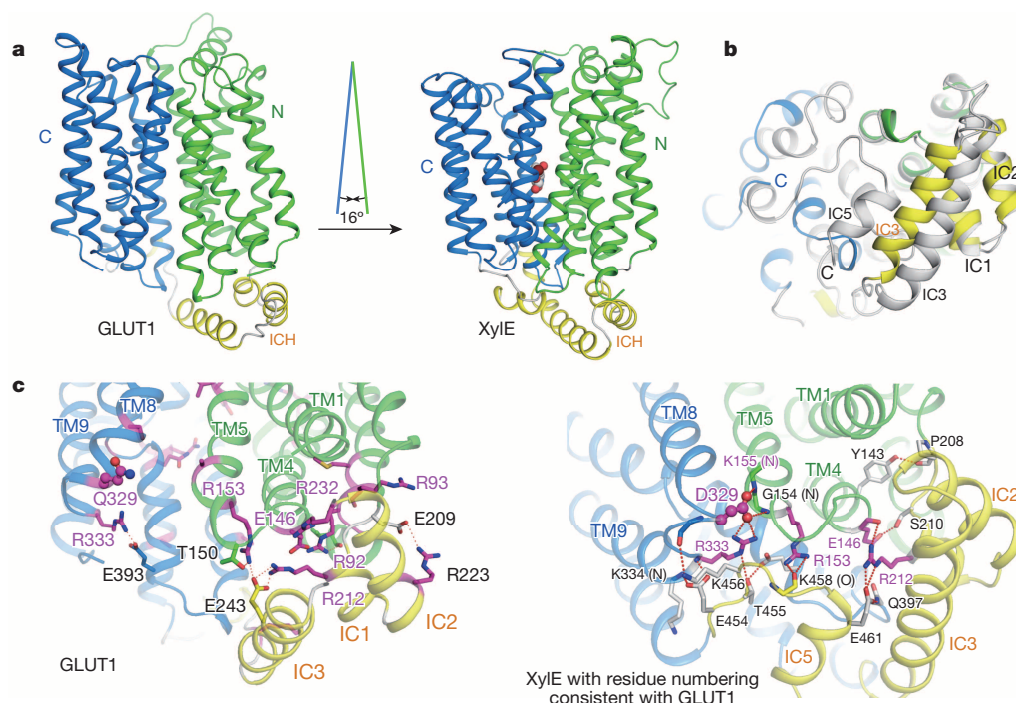
In a simplified model, alternating access of the MFS transporters can be achieved through rigid-body rotation of the N and C domains<sup>25,46</sup>. Structural comparison of the inward-open GLUT1 and the outward-facing XyleE reveals an approximately 16° concentric rotation of the two domains (Fig. 3a). GLUT1 and XyleE share three intracellular helices (Extended Data Fig. 4a). Intracellular helix 1 (IC1) and IC2 appear to maintain a defined conformation with respect to the N domain, whereas IC3 is pulled towards the C domain during the inter-domain rotation from outward-facing to inward-open (Fig. 3b and Extended Data Fig. 4b, c). To facilitate structural comparison, we will use the GLUT1 residue numbers to annotate the conserved residues in XyleE (Fig. 3c and Extended Data Fig. 5).

Nearly half of the mutations derived from GLUT1 deficiency syndrome map to residues that mediate the inter-domain contacts on the cytosolic side (Fig. 3c and Extended Data Fig. 5). In the outward-facing XyleE, several conserved GLUT1-deficiency-syndrome-associated residues contribute to the interactions between the N and C domains along the intracellular side of the membrane, thereby stabilizing the outward-facing conformation (Figs 2 and 3c). The guanidinium group of Arg 153 at the intracellular end of transmembrane 5 (TM5) forms two hydrogen bonds with the carbonyl oxygen of Lys 458 which is on the loop preceding IC5. Arg 212 on IC2 interacts with Glu 461 on IC5. Most notably, Asp 329 seems to have a critical role in the maintenance of the outward-facing conformation of XyleE with its carboxylate group accepting two



**Figure 2 | Structural mapping of disease-derived mutations in GLUT1.** The residues for which mutations have been identified in patients with GLUT1 deficiency syndrome and other symptoms are coloured magenta. The orange circle and the grey and cyan shades indicate the three major clusters of residues for which mutations are associated with diseases. The N, C and ICH domains are coloured green, blue and yellow, respectively. The same colour scheme is used in all subsequent figures. Details of the mutations can be found in Extended Data Table 2.





**Figure 3 | The ICH domain serves as a latch that tightens the intracellular gate.** **a**, Structural comparison of the inward-open GLUT1 to its *E. coli* homologue Xyle (PDB accession code 4GC0)<sup>28</sup>. **b**, Structural overlay of GLUT1 and Xyle using their respective N domains. For visual clarity, GLUT1 is displayed in three colours and Xyle is coloured grey. **c**, Analysis of the disease-related GLUT1 residues that mediate the inter-domain interactions on the intracellular side. Left: positioning and interactions of the disease-related residues on the intracellular side of the inward-open GLUT1. Right: the interaction network of the corresponding residues in the outward-facing Xyle. The residues shown in this panel are conserved between GLUT1 and Xyle, with the disease-related residues coloured and labelled magenta. The Xyle residues are numbered according to their counterparts in GLUT1. Hydrogen bonds are represented by dashed lines.

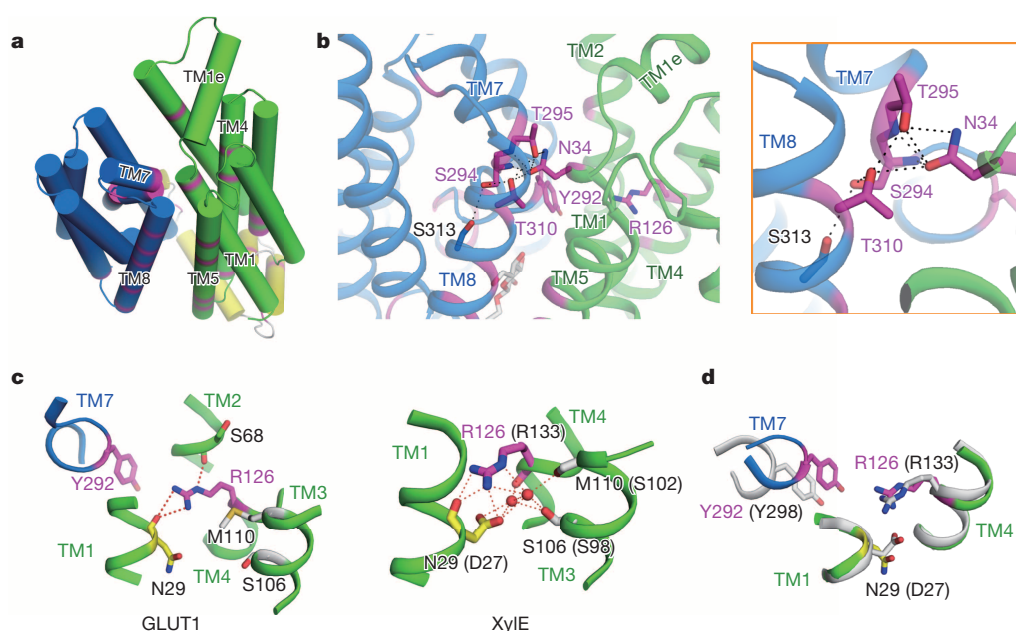
hydrogen bonds from the backbone amide groups of Gly 154 and Lys 155. These hydrogen bonds are further stabilized by a pair of charge-stabilized hydrogen bonds between Asp 329 and Arg 333.

Obviously, none of these interactions observed in Xyle would exist in the inward-open conformation of GLUT1 (Fig. 3c). Asp 329 in Xyle is replaced by Glu in GLUT1. Assuming a similar set of interactions in GLUT1 as in Xyle, the mutation E329Q is predicted to weaken or abolish the hydrogen bonds with the amide groups of Gly 154 and Ala 155. Consequently, GLUT1(E329Q) may favour the inward-facing conformation as it probably encounters an increased energy barrier during the outward transition. Along this line, we predict that substitution of other residues in this interface may weaken the interactions between the N and C domains on the intracellular side, thus impeding the transport cycle by disfavours the outward-facing conformation.

Analysis of the cluster II mutations suggests that the ICH domain, which represents a unique feature of the sugar porter subfamily, may function as a latch to secure closure of the intracellular gate in the outward-facing conformation. Supporting this notion, the interactions between the N and C domains in the transmembrane region of the outward-facing Xyle are mediated by a limited number of hydrophobic residues (Extended Data Fig. 3b). Therefore, the inter-domain contacts involving the charged or polar ICH residues are likely to have a critical role to close the transporter on the intracellular side during an alternating access cycle.

### The extracellular gate

In the inward-open conformation of GLUT1, TM1 and TM7 contact each other on the extracellular side, representing the major constituents of



**Figure 4 | Analysis of the extracellular gate of the inward-open GLUT1.** **a**, TM1 and TM7 represent the major constituents of the extracellular gate in the inward-open GLUT1. **b**, Several disease-related residues mediate the interactions between the N and C domains on the extracellular side of the inward-open GLUT1. A close-up view of the hydrogen bond network is shown in the inset. **c**, Comparison of the local interactions for the disease-related residue Arg 126 in GLUT1 and the corresponding residue Arg 133 in Xyle. The residues of Xyle are labelled according to their counterparts in GLUT1 whereas their native numbers are indicated in brackets. **d**, Arg 126 on TM4 of GLUT1 may form a cation- $\pi$  interaction with Tyr 292 on TM7 of the C domain, contributing to the affinity of the extracellular gate. The structures of GLUT1 and Xyle are superimposed using their respective N domains.

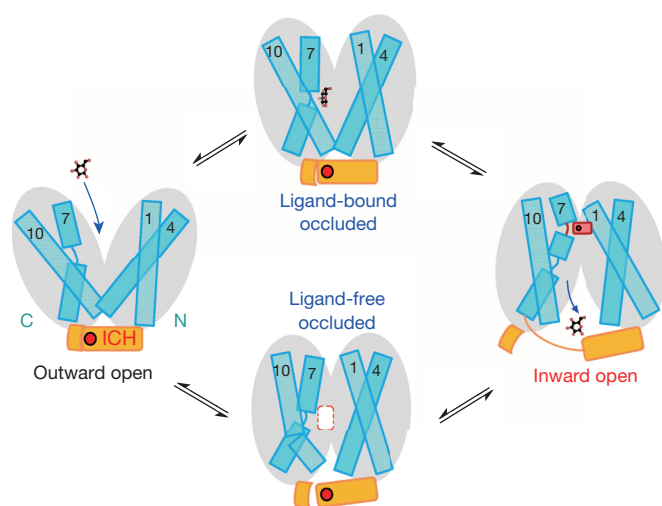


the extracellular gate (Fig. 4a). It is noteworthy that TM7 of GLUT1 contains one extra kink close to its extracellular end compared to that of the outward-facing XylE (Extended Data Fig. 4d). Notably, the mutations N34I, N34S, N34Y, S294P, T295M and T310I, affecting four amino acids, were isolated from GLUT1 deficiency syndrome patients (Extended Data Table 2). Asn 34 on TM1 appears to be an organizing centre that mediates hydrogen bonds to Ser 294 and Thr 295 on TM7 and Thr 310 on TM8 (Fig. 4b). Together, this network of hydrogen bonds between the N and C domains occludes the ligand-binding site from the extracellular environment in the inward-facing conformation of GLUT1. Note that the residues Tyr 292, Ser 294 and Thr 295, which interact with the N domain, are located surrounding the discontinuous helical segment of TM7 (Fig. 4b).

Among the disease-related residues, Arg 126 is of particular interest, because its counterpart in XylE, Arg 133, forms a network of hydrogen bonds with Asp 27, an essential residue for proton coupling in the outward-facing XylE (ref. 45). Asp 27 in XylE is replaced by Asn 29 in GLUT1. Notably, Arg 126 does not interact with the side chain of Asn 29 in the inward-open GLUT1 (Fig. 4c). More importantly, the guanidinium group of Arg 126 is approximately 6 Å away from the benzene ring of Tyr 292 on TM7, probably forming a cation- $\pi$  interaction that further strengthens the extracellular gate in the inward-open GLUT1 (Fig. 4d).

## Discussion

Structural resolution of the human glucose transporter GLUT1 serves as a framework for understanding its functional mechanism. Comparison of the proton-independent uniporter GLUT1 and its bacterial homologue, the proton symporter XylE, the structures of which were obtained in distinct conformations<sup>28,29</sup>, allows mechanistic analysis of facilitative diffusion versus active transport. Conclusions derived from such analysis may be generally applicable to other uniporters and proton-coupled symporters. On the basis of our structural analysis and published biochemical data, we propose a working model for GLUT1 (Fig. 5).



**Figure 5 | A working model for GLUT1.** Shown here are the predicted conformations—outward-open, ligand-bound and occluded, inward-open, and ligand-free and occluded—required for a complete transport cycle according to the alternating access model. The inward-open conformation of GLUT1 is reported in this study. The ligand-bound, occluded conformation and the ligand-free, occluded one are predicted from two XylE structures in the outward-facing, partly occluded and ligand-bound state<sup>28</sup>, and the inward-facing, occluded state<sup>29</sup>, respectively. The outward-open structure remains to be captured. The ICH domain is illustrated as a latch that strengthens the intracellular gate in the outward-facing conformations. The extracellular gate comprises a few residues from TM1, TM4 and TM7 that are illustrated by the red brick in the ‘inward open’ cartoon.

Uniporters only catalyse the translocation of substrate down its concentration gradient. The essential elements for a uniporter to complete a transport cycle are the conformational switches. In the case of GLUT1, because of the extensive interactions between TMD and ICH, the ligand-free protein may prefer an outward-open conformation. Substrate binding at the primary site on the C domain may provide extra contacts with the N domain, inducing closure of the N and C domains on the extracellular side and leading to rearrangement of interactions on both sides of the bound substrate. When the binding affinity between the N and C domains on the extracellular side (Fig. 4) surpasses that of the intracellular side (Fig. 3c), the protein may switch to the inward-open conformation. Once GLUT1 adopts the inward-open conformation, the substrate is exposed to a low concentration environment; then the equilibrium is shifted towards substrate dissociation. The substrate-free uniporter is able to return to the outward-facing state. This model predicts that substrate-free uniporters have a preferred open conformation, and substrate association and dissociation drive the conformational switches.

Proton symporters exploit the transmembrane proton gradient (also known as the proton motive force<sup>47</sup>) to drive the ‘uphill’ translocation of substrate against its concentration gradient. The translocation of proton and substrate are obligatorily coupled. Notably, Asp 27 of XylE, corresponding to Asp 32 in GalP (ref. 48) and Asp 22 in GlcP (ref. 30), has a critical role in proton coupling. The mutation D27N in XylE or D22N in GlcP led to abrogation of proton-dependent active symport, but not the counterflow activity<sup>30,45</sup>. The structural consequence of protonation of Asp is thought to be similar to Asn substitution. Therefore the structure of GLUT1, where Asn 29 is the counterpart of the critical Asp in the bacterial homologues, provides an opportunity to understand the protonated state of the proton symporters.

Protonation of the key Asp residue releases the counterpart of Arg 126 in GLUT1 which is subsequently able to form a cation- $\pi$  interaction with the aromatic residue in the C domain, an event that probably triggers the outward to inward transition (Fig. 4c, d). In this partial process, a uniporter can be regarded as a permanently protonated symporter. However, unlike uniporters which upon substrate release are able to return from inward- to outward-facing, a substrate-released symporter cannot undergo the conformational switch until it is deprotonated, and vice versa (that is, deprotonated symporter cannot go back to outward-facing unless the substrate is dissociated). Detailed mechanistic understanding of the transport mechanisms of uniporters and symporters awaits additional biochemical and biophysical investigations. In particular, the coupling mechanism of deprotonation, substrate release into a high-concentration environment, and the conformational switch from inward-facing to outward-facing of symporters remains to be elucidated.

In summary, we report the long-sought-after crystal structure of the human glucose transporter GLUT1, which allows accurate mapping and potential mechanistic interpretation of disease-related mutations. Structural comparison of GLUT1 with XylE provides a foundation to understand the transport mechanisms of facilitative uniporters and proton-driven symporters. The structures of GLUT1, XylE and GlcP provide important insight into structural and mechanistic understanding of the sugar porter family whose members are involved in a wide variety of biological processes as well as biotechnological applications<sup>49</sup>. The structure also serves as a guiding principle for the development of potential therapeutic agents that target GLUT1 and other physiologically important MFS sugar transporters.

## METHODS SUMMARY

To obtain a large quantity of GLUT1 protein suitable for crystallization trials, we exploited multiple recombinant expression systems, including *E. coli*, several yeast species, baculovirus-infected insect cells, and mammalian expression systems. On the basis of both protein yield and solution behaviour, insect cells were chosen for large-scale production of GLUT1. Treatment of the purified protein by deglycosylase resulted in shift to a lower band on SDS-PAGE by western blot, indicating post-translational glycosylation of GLUT1 during expression. To eliminate potential heterogeneity caused by glycosylation, we introduced a single point mutation,

N45T, in GLUT1. A second point mutation, E329Q, was introduced in the hope of stabilizing GLUT1 in a certain conformation. The full-length human GLUT1 variant (residues 1–492, N45T and E329Q) with a C-terminal 10× His tag was overexpressed in the High Five insect cells infected with baculoviruses. Forty-eight hours after viral infection, the cells were collected and disrupted using dounce homogenizer. The protein was extracted from membrane fraction with 2% (w/v) *n*-dodecyl- $\beta$ -D-maltoside (DDM) at 4 °C for 2 h and purified with nickel affinity resin (Ni-NTA) in the presence of 0.05% (w/v) DDM and 5% glycerol (w/v). The protein eluted from Ni-NTA was concentrated to about 10 mg ml<sup>-1</sup> before applying to size-exclusion chromatography (Superdex-200; GE Healthcare) in the buffer containing 25 mM MES pH 6.0, 150 mM NaCl, 5% glycerol, and 0.4% (w/v)  $\beta$ -NG. The peak fractions were collected for crystallization trials. Crystals appeared through the hanging-drop vapour diffusion method at 4 °C in the well buffer containing 30% (w/v) PEG400, 0.1 M MES pH 6.0, and 0.1 M MgCl<sub>2</sub>. Data sets were collected at Shanghai Synchrotron Radiation Facility (SSRF) beamline BL17U. The structure of GLUT1 was determined using molecular replacement with the coordinates of the N and C domains of Xyle as separate searching models.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 January; accepted 1 April 2014.

Published online 18 May 2014.

- Mueckler, M. *et al.* Sequence and structure of a human glucose transporter. *Science* **229**, 941–945 (1985).
- Thorens, B. & Mueckler, M. Glucose transporters in the 21st Century. *Am. J. Physiol. Endocrinol. Metab.* **298**, E141–E145 (2010).
- Kasahara, M. & Hinkle, P. C. Reconstitution and purification of the D-glucose transporter from human erythrocytes. *J. Biol. Chem.* **252**, 7384–7390 (1977).
- Dick, A. P., Harik, S. I., Klip, A. & Walker, D. M. Identification and characterization of the glucose transporter of the blood-brain barrier by cytochalasin B binding and immunological reactivity. *Proc. Natl Acad. Sci. USA* **81**, 7233–7237 (1984).
- Pardridge, W. M., Boado, R. J. & Farrell, C. R. Brain-type glucose transporter (GLUT-1) is selectively localized to the blood-brain barrier. Studies with quantitative western blotting and in situ hybridization. *J. Biol. Chem.* **265**, 18035–18040 (1990).
- Maher, F., Vannucci, S. J. & Simpson, I. A. Glucose transporter proteins in brain. *FASEB J.* **8**, 1003–1011 (1994).
- Klepper, J. *et al.* Defective glucose transport across brain tissue barriers: a newly recognized neurological syndrome. *Neurochem. Res.* **24**, 587–594 (1999).
- Brockmann, K. The expanding phenotype of GLUT1-deficiency syndrome. *Brain Dev.* **31**, 545–552 (2009).
- Pearson, T. S., Akman, C., Hinton, V. J., Engelstad, K. & De Vivo, D. C. Phenotypic spectrum of glucose transporter type 1 deficiency syndrome (Glut1 DS). *Curr. Neurol. Neurosci. Rep.* **13**, 342 (2013).
- Saudubray, J.-M., van den Berghe, G. & Walter, J. H. *Inborn Metabolic Diseases: Diagnosis and Treatment* 5th edn 656 (Springer, 2011).
- Leen, W. G. *et al.* GLUT1 deficiency syndrome into adulthood: a follow-up study. *J. Neurol.* **261**, 589–599 (2014).
- Suls, A. *et al.* Early-onset absence epilepsy caused by mutations in the glucose transporter GLUT1. *Ann. Neurol.* **66**, 415–419 (2009).
- Amann, T. & Hellerbrand, C. GLUT1 as a therapeutic target in hepatocellular carcinoma. *Expert Opin. Ther. Targets* **13**, 1411–1427 (2009).
- Amann, T., Kirovski, G., Bosserhoff, A. K. & Hellerbrand, C. Analysis of a promoter polymorphism of the GLUT1 gene in patients with hepatocellular carcinoma. *Mol. Membr. Biol.* **28**, 182–186 (2011).
- Amann, T. *et al.* GLUT1 expression is increased in hepatocellular carcinoma and promotes tumorigenesis. *Am. J. Pathol.* **174**, 1544–1552 (2009).
- Shim, B. Y. *et al.* Glucose transporter 1 (GLUT1) of anaerobic glycolysis as predictive and prognostic values in neoadjuvant chemoradiotherapy and laparoscopic surgery for locally advanced rectal cancer. *Int. J. Colorectal Dis.* **28**, 375–383 (2013).
- Ramani, P., Headford, A. & May, M. T. GLUT1 protein expression correlates with unfavourable histologic category and high risk in patients with neuroblastic tumours. *Virchows Arch.* **462**, 203–209 (2013).
- McGuire, B. B. & Fitzpatrick, J. M. Biomarkers in renal cell carcinoma. *Curr. Opin. Urol.* **19**, 441–446 (2009).
- Kaira, K. *et al.* Biological significance of F-FDG uptake on PET in patients with non-small-cell lung cancer. *Lung Cancer* **83**, 197–204 (2014).
- Maiden, M. C., Davis, E. O., Baldwin, S. A., Moore, D. C. & Henderson, P. J. Mammalian and bacterial sugar transport proteins are homologous. *Nature* **325**, 641–643 (1987).
- Hediger, M. A., Clemençon, B., Burrier, R. E. & Bruford, E. A. The ABCs of membrane transporters in health and disease (SLC series): introduction. *Mol. Aspects Med.* **34**, 95–107 (2013).
- Shi, Y. Common folds and transport mechanisms of secondary active transporters. *Annu. Rev. Biophys.* **42**, 51–72 (2013).
- Henderson, P. J. & Baldwin, S. A. This is about the in and the out. *Nature Struct. Mol. Biol.* **20**, 654–655 (2013).
- Pao, S. S., Paulsen, I. T. & Saier, M. H. Jr. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62**, 1–34 (1998).
- Yan, N. Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem. Sci.* **38**, 151–159 (2013).
- Radestock, S. & Forrest, L. R. The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *J. Mol. Biol.* **407**, 698–715 (2011).
- Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969–970 (1966).
- Sun, L. *et al.* Crystal structure of a bacterial homologue of glucose transporters GLUT1–4. *Nature* **490**, 361–366 (2012).
- Quistgaard, E. M., Low, C., Moberg, P., Tresaugues, L. & Nordlund, P. Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. *Nature Struct. Mol. Biol.* **20**, 766–768 (2013).
- Iancu, C. V., Zamoon, J., Woo, S. B., Aleshin, A. & Choe, J. Y. Crystal structure of a glucose/H<sup>+</sup> symporter and its mechanism of action. *Proc. Natl Acad. Sci. USA* **110**, 17862–17867 (2013).
- Nelson, D. L. & Cox, M. M. *Lehninger Principles of Biochemistry* (W. H. Freeman, 2008).
- Schürmann, A. *et al.* Role of conserved arginine and glutamate residues on the cytosolic surface of glucose transporters for transporter function. *Biochemistry* **36**, 12897–12902 (1997).
- Smirnova, I., Kasho, V. & Kaback, H. R. Lactose permease and the alternating access mechanism. *Biochemistry* **50**, 9684–9693 (2011).
- Wang, D., Kranz-Eble, P. & De Vivo, D. C. Mutational analysis of GLUT1 (SLC2A1) in Glut-1 deficiency syndrome. *Hum. Mutat.* **16**, 224–231 (2000).
- Klepper, J. *et al.* Autosomal dominant transmission of GLUT1 deficiency. *Hum. Mol. Genet.* **10**, 63–68 (2001).
- Overweg-Plandsoen, W. C. *et al.* GLUT-1 deficiency without epilepsy—an exceptional case. *J. Inher. Metab. Dis.* **26**, 559–563 (2003).
- Wang, D. *et al.* Glut-1 deficiency syndrome: clinical, genetic, and therapeutic aspects. *Ann. Neurol.* **57**, 111–118 (2005).
- Weber, Y. G. *et al.* GLUT1 mutations are a cause of paroxysmal exertion-induced dyskinesias and induce hemolytic anemia by a cation leak. *J. Clin. Invest.* **118**, 2157–2168 (2008).
- Schneider, S. A. *et al.* GLUT1 gene mutations cause sporadic paroxysmal exercise-induced dyskinesias. *Mov. Disord.* **24**, 1684–1688 (2009).
- Leen, W. G. *et al.* Glucose transporter-1 deficiency syndrome: the expanding clinical and genetic spectrum of a treatable disorder. *Brain* **133**, 655–670 (2010).
- Mullen, S. A., Suls, A., De Jonghe, P., Berkovic, S. F. & Scheffer, I. E. Absence epilepsies with widely variable onset are a key feature of familial GLUT1 deficiency. *Neurology* **75**, 432–440 (2010).
- Urbizu, A. *et al.* Paroxysmal exercise-induced dyskinesia, writer's cramp, migraine with aura and absence epilepsy in twin brothers with a novel SLC2A1 missense mutation. *J. Neurol. Sci.* **295**, 110–113 (2010).
- Anheim, M. *et al.* Excellent response to acetazolamide in a case of paroxysmal dyskinesias due to GLUT1-deficiency. *J. Neurol.* **258**, 316–317 (2011).
- Striano, P. *et al.* GLUT1 mutations are a rare cause of familial idiopathic generalized epilepsy. *Neurology* **78**, 557–562 (2012).
- Madej, M. G., Sun, L., Yan, N. & Kaback, H. R. Functional architecture of MFS D-glucose transporters. *Proc. Natl Acad. Sci. USA* **111**, E719–E727 (2014).
- Dang, S. *et al.* Structure of a fucose transporter in an outward-open conformation. *Nature* **467**, 734–738 (2010).
- Mitchell, P. David Keilin's respiratory chain concept and its chemiosmotic consequences. In *Nobel Lectures, Chemistry 1971–1980* (ed. Forsén, S.) (World Scientific Publishing Co., 1978).
- Sanderson, N. M., Qi, D., Steel, A. & Henderson, P. J. Effect of the D32N and N300F mutations on the activity of the bacterial sugar transport protein, GalP. *Biochem. Soc. Trans.* **26**, S306 (1998).
- Henderson, P. J. & Baldwin, S. A. Structural biology: Bundles of insights into sugar transporters. *Nature* **490**, 348–350 (2012).
- DeLano, W. L. The PyMOL Molecular Graphics System. <http://www.pymol.org> (2002).

**Acknowledgements** We thank J. He, L. Tang, F. Yu and S. Huang at Shanghai Synchrotron Radiation Facility (SSRF). This work was supported by funds from the Ministry of Science and Technology (grant number 2011CB910501), Projects 31321062-20131319400, 31125009 and 91017011 of the National Natural Science Foundation of China, and funds from Tsinghua-Peking Center for Life Sciences. The research of N.Y. was supported in part by an International Early Career Scientist grant from the Howard Hughes Medical Institute.

**Author Contributions** N.Y. conceived the project. D.D. and N.Y. designed all experiments. D.D., C.X., P.S., J.W., C.Y. and M.H. performed the experiments. All authors analysed the data and contributed to manuscript preparation. N.Y. wrote the manuscript.

**Author Information** The X-ray crystallographic coordinates and structure factor files of human GLUT1(N45T/E329Q) have been deposited in the Protein Data Bank (PDB) with the accession code 4PYP. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.Y. ([nyan@tsinghua.edu.cn](mailto:nyan@tsinghua.edu.cn)).

## METHODS

**Protein expression and purification.** The full-length human GLUT1 complementary DNA was subcloned into the NdeI and XhoI sites of pFastBac1 (Invitrogen) with a C-terminal 10× His tag. All GLUT1 mutants were generated with a standard PCR-based strategy. The recombinant GLUT1 was expressed using the pFastBac baculovirus system (Invitrogen). Briefly, bacmid DNAs were generated in DH10Bac cells, and the resulting baculoviruses were generated and amplified in Sf9 insect cells (Invitrogen). GLUT1 was overexpressed in High Five insect cells (Invitrogen) grown in the SIM HF medium (Sino Biological Inc.). Forty-eight hours after viral infection, the cells were collected and homogenized in the buffer containing 25 mM Tris pH 8.0 and 150 mM NaCl. The cells were disrupted using dounce homogenizer for 80 cycles on ice. Cell debris was removed by low-speed centrifugation for 10 min. The supernatant was applied to ultracentrifugation at 150,000g for 1 h. The membrane fraction was solubilized in TBS buffer (25 mM Tris pH 8.0 and 150 mM NaCl) containing protease inhibitors (aprotinin at 0.8 μM, pepstatin at 2 μM, leupeptin at 5 μg ml<sup>-1</sup>) and 2% (w/v) *n*-dodecyl-β-D-maltoside (DDM, Anatrace) at 4 °C for 2 h. After a further step of ultracentrifugation at 150,000g for 30 min, the detergent-soluble fraction was harvested and incubated with nickel affinity resin (Ni-NTA, Qiagen) at 4 °C for 30 min. The resin was rinsed with the buffer containing 25 mM MES pH 6.0, 150 mM NaCl, 30 mM imidazole, 5% glycerol (w/v), and 0.05% (w/v) DDM for three times. The protein was eluted from the affinity resin with the wash buffer plus 300 mM imidazole. The protein was concentrated to about 10 mg ml<sup>-1</sup> before further purification by size-exclusion chromatography (Superdex-200; GE Healthcare) in the buffer containing 25 mM MES pH 6.0, 150 mM NaCl, 5% glycerol, and selected detergents. The peak fractions were collected and flash frozen in liquid nitrogen for crystallization.

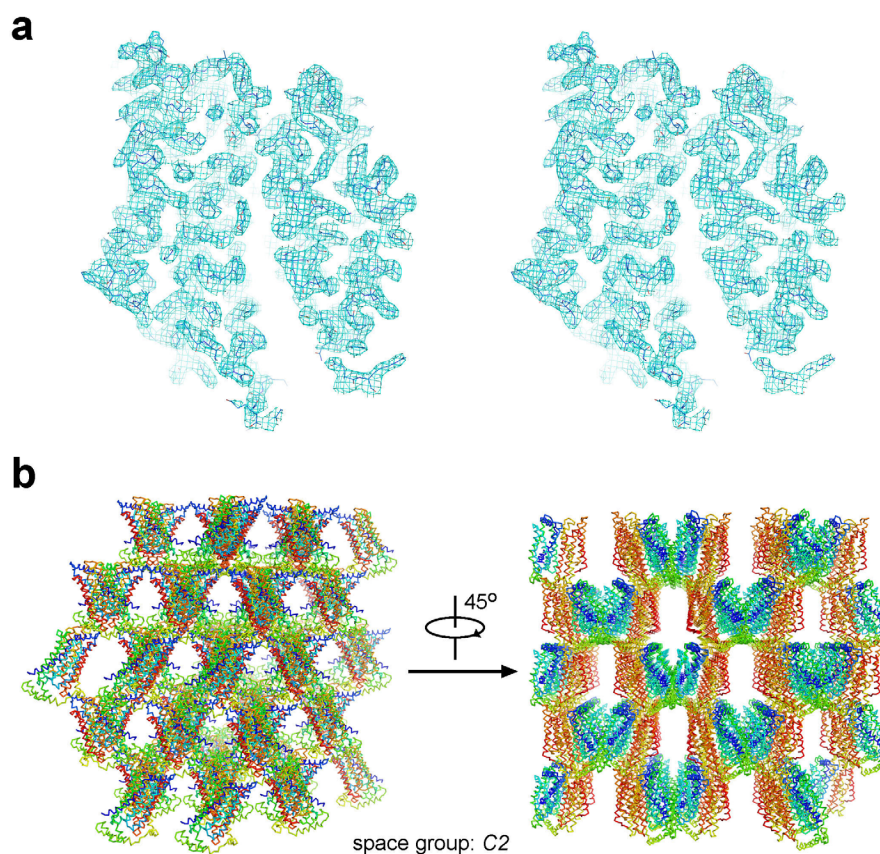
**Crystallization.** Extensive crystallization trials were performed for GLUT1 proteins purified in various detergents and in lipidic cubic phase. Most of the efforts failed to yield crystals. Finally, the full-length GLUT1 containing two point mutations N45T and E329Q purified in the presence of 0.4% (w/v) β-NG gave rise to crystals through hanging-drop vapour diffusion method in the well buffer containing 30% (w/v) PEG400, 0.1 M MES pH 6.0 and 0.1 M MgCl<sub>2</sub> at 4 °C. The crystals

appeared after 2 days and reached full size in 5 to 7 days. A few crystals diffracted X-rays to beyond 3.2 Å at Shanghai Synchrotron Radiation Facility (SSRF) beamline BL17U.

**Data collection and structure determination.** All data sets were collected at SSRF beamline BL17U and processed with the HKL2000 packages<sup>51</sup>. Further processing was carried out with programs from the CCP4 suite<sup>52</sup>. Data collection and structure refinement statistics are summarized in Extended Data Table 1. The phase was solved by molecular replacement using PHASER<sup>53</sup> with Xyle (PDB code 4GBY) as searching model. The model was first modified by CHAINSAW<sup>54</sup>, then the N domain and C domain were taken separately as input ensembles for PHASER. The model was further rebuilt in COOT<sup>55</sup> and refined with PHENIX<sup>56</sup>. The sequence docking was further aided by sequence alignment with Xyle.

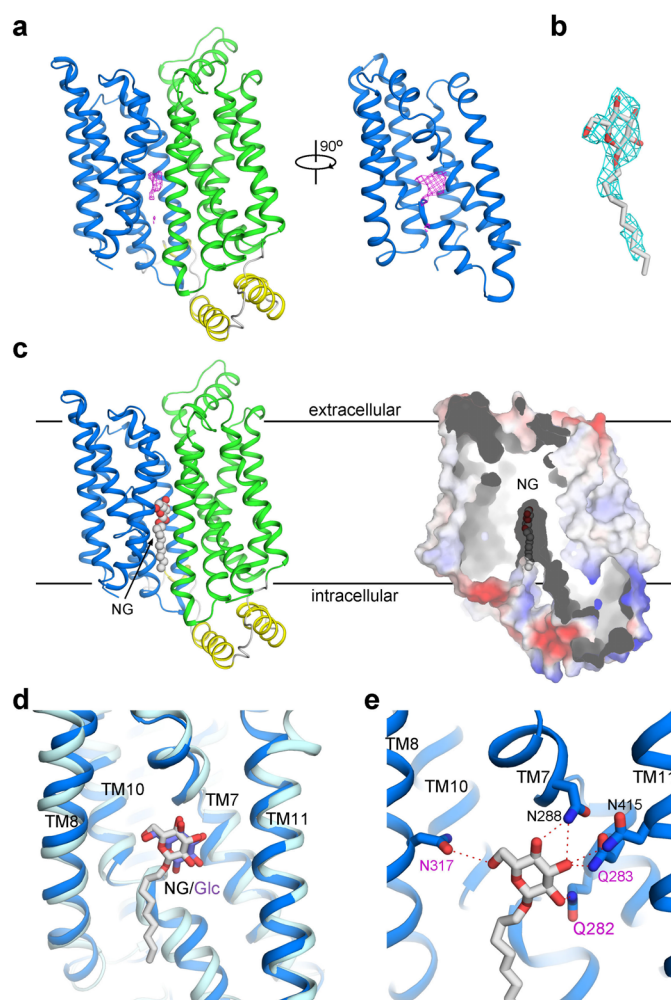
51. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data. *Methods Enzymol.* **276**, 307–326 (1997).
52. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
53. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
54. Stein, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641–643 (2008).
55. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
56. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
57. Pérez-Dueñas, B. *et al.* Childhood chorea with cerebral hypotrophy: a treatable GLUT1 energy failure syndrome. *Arch. Neurol.* **66**, 1410 (2009).
58. Pascual, J. M., Van Heertum, R. L., Wang, D., Engelstad, K. & De Vivo, D. C. Imaging the metabolic footprint of Glut1 deficiency on the brain. *Ann. Neurol.* **52**, 458–464 (2002).
59. Afawi, Z. *et al.* Mild adolescent/adult onset epilepsy and paroxysmal exercise-induced dyskinesia due to GLUT1 deficiency. *Epilepsia* **51**, 2466–2469 (2010).
60. Klepper, J. *et al.* Autosomal recessive inheritance of GLUT1 deficiency syndrome. *Neuropediatrics* **40**, 207–210 (2009).





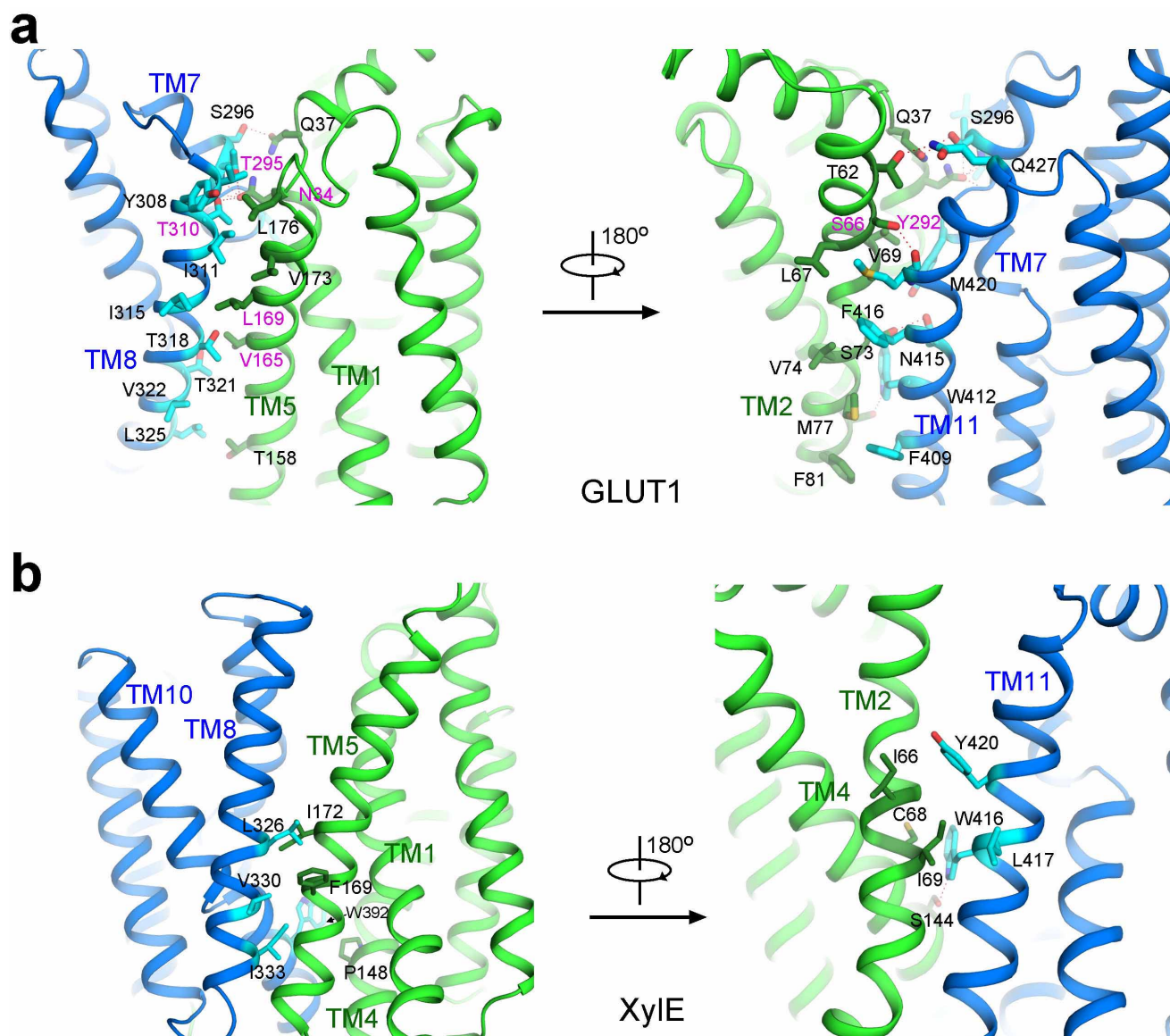
**Extended Data Figure 1 | Structure determination of GLUT1.** **a**, The  $2F_o - F_c$  electron density map. The stereo-view map for one representative slab, shown as cyan mesh, is contoured at  $1.0\sigma$ . **b**, The crystal packing of GLUT1 in

the space group C2. Each GLUT1 molecule is shown as rainbow-coloured ribbon, blue and red for the N and C termini, respectively.



**Extended Data Figure 2 | One  $\beta$ -NG molecule occupies the substrate-binding site of the inward-open GLUT1.** **a**, The 'omit' electron density observed in the inward-open cavity of GLUT1. The electron density, shown as magenta mesh, is contoured at  $3.0\sigma$ . The N, C and ICH domains are coloured green, blue and yellow, respectively. **b**, A  $\beta$ -NG molecule fits well into the electron density inside the cavity. The  $2F_o - F_c$  electron density map (cyan mesh) for the  $\beta$ -NG molecule is contoured at  $1.0\sigma$ . **c**, The overall GLUT1 structure with the bound  $\beta$ -NG molecule.  $\beta$ -NG is represented by white spheres. **d**, The coordination of the sugar moiety of  $\beta$ -NG by GLUT1 is similar to the binding of D-glucose by XylE. The structures of GLUT1 (blue) and XylE

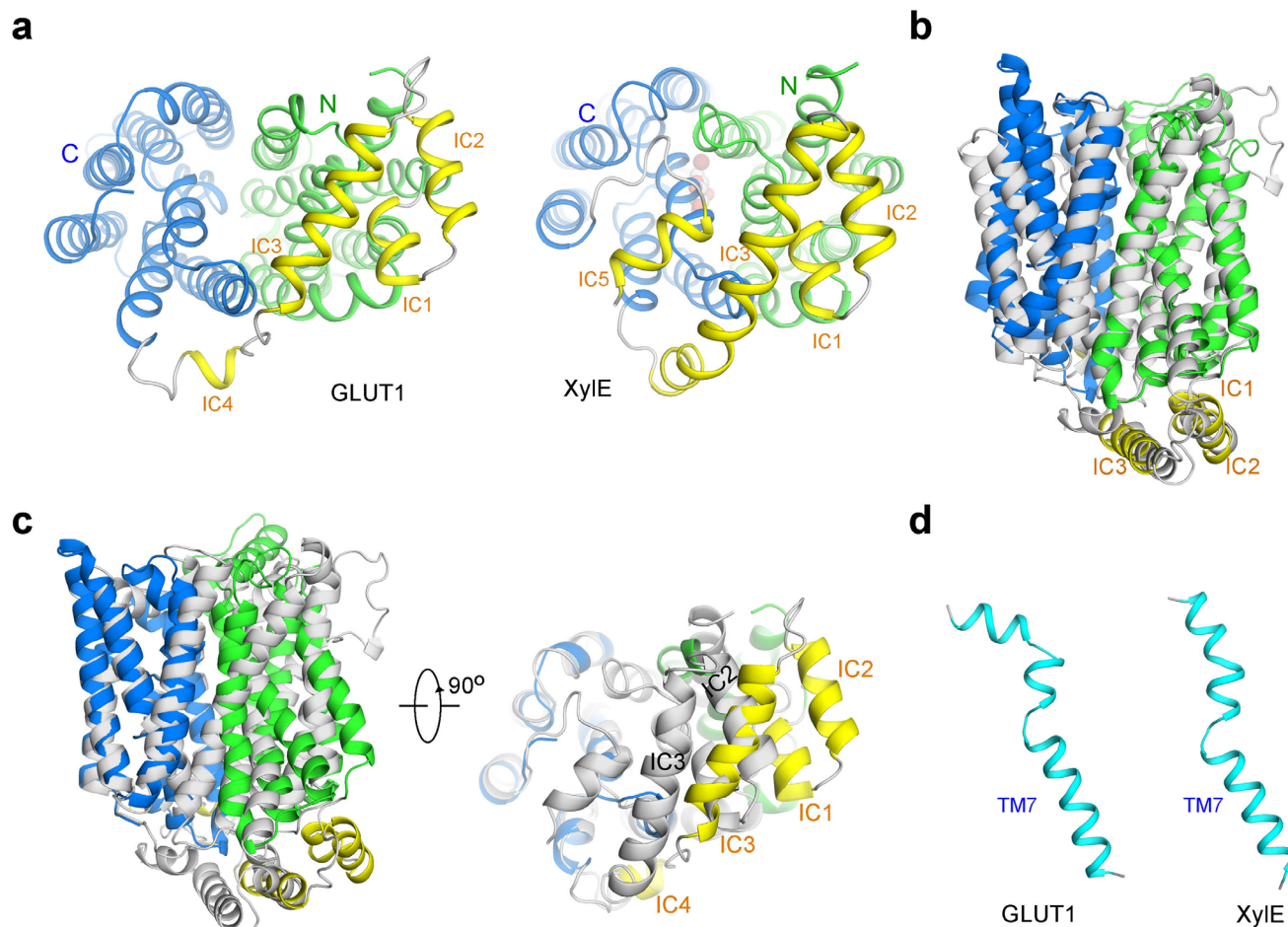
(cyan) are superimposed relative to their respective C domains. The ligands are shown in stick representation. Despite the similarity between D-glucose and  $\beta$ -NG, we cannot exclude the possibility that presence of the aliphatic tail of  $\beta$ -NG in GLUT1 may subtly affect positioning of the sugar moiety compared to D-glucose. **e**, Coordination of the D-glucopyranoside of  $\beta$ -NG by GLUT1. The D-glucopyranoside of  $\beta$ -NG is hydrogen-bonded to the surrounding polar residues in the C domain, including Gln 282/Gln 283/Asn 288 from TM7, Asn 317 from TM8, and Asn 415 from TM11. The residues whose mutations are associated with GLUT1 deficiency syndrome are labelled in magenta.



**Extended Data Figure 3 | Interactions between the N and C domains observed in the inward-open GLUT1 and the outward-facing XylE.** a, In GLUT1, the inter-domain contacts, mainly on the extracellular side, include both van der Waals interactions and hydrogen bonds. Residues whose

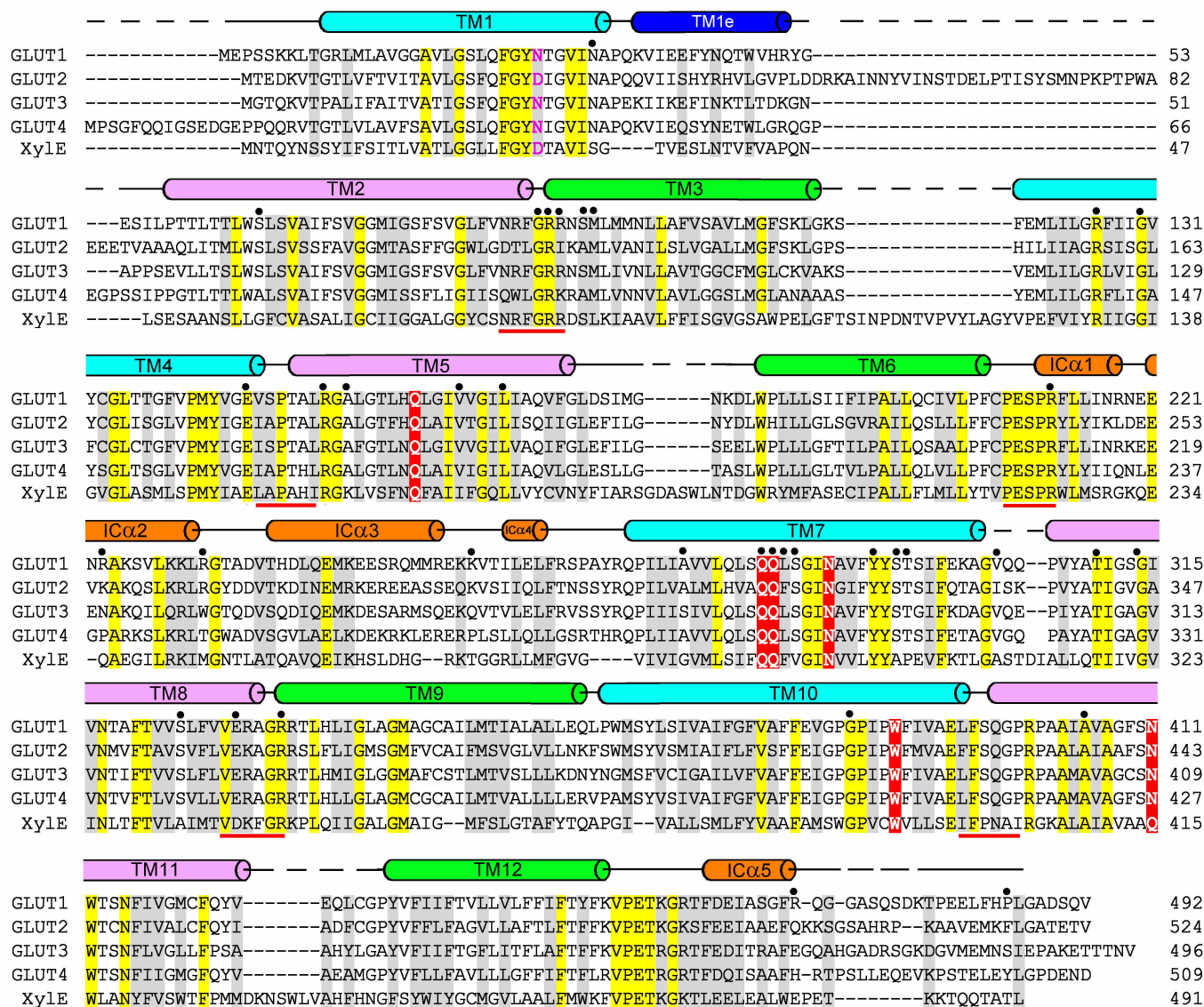
mutations are associated with GLUT1 deficiency syndrome are labelled in magenta. b, In XylE, there are only limited interactions between the N and C domains at the transmembrane region.





**Extended Data Figure 4 | Conformational differences between GLUT1 and XylE.** **a**, Structural comparison of the inward-open GLUT1 and the outward-facing, partly occluded, ligand-bound XylE. Similar intracellular views are shown for GLUT1 and XylE. Note that the C-terminal helix IC5, which was referred to as IC4 in a previous study of XylE (ref. 28), is invisible in the structure of GLUT1 probably due to its inherent flexibility in this

conformation. **b**, **c**, Structural superimpositions of GLUT1 and XylE relative to their respective N domains (**b**) and C domains (**c**). A detailed analysis can be found in Figs 3 and 4. **d**, Conformational differences of TM7 between GLUT1 and XylE. Compared to that in XylE, the extracellular segment of TM7 in GLUT1 is further bent away from the transport path.



**Extended Data Figure 5 | Sequence alignment of GLUT1-4 with XylE.**

Secondary structural elements of GLUT1 are indicated above the sequence alignment. Invariant and highly conserved amino acids are shaded yellow and grey, respectively. The conserved sugar porter family signature motifs are

underscored with red lines. The residues that are hydrogen-bonded to D-glucose in XylE are shaded red. The GLUT1 residues whose mutations were found in GLUT1 deficiency syndrome are indicated by black circles above. The sequences were aligned with ClustalW.

Extended Data Table 1 | Statistics of data collection and refinement for GLUT1

Protein	GLUT1
Integration Package	HKL2000
Space Group	C2
Unit Cell (Å)	120.51, 102.00, 65.60
Unit Cell (°)	90, 101.27, 90
Wavelength (Å)	0.9791
Resolution (Å)	40~3.20 (3.31~3.20)
R <sub>merge</sub> (%)	10.2 (53.8)
I/sigma	17.6 (3.2)
Completeness (%)	98.2 (97.2)
Number of measured reflections	52,880
Number of unique reflections	12,834
Redundancy	4.1 (4.2)
R <sub>work</sub> / R <sub>free</sub> (%)	24.14/27.35
No. atoms	
Protein	3,469
main chain	1,788
side chain	1,681
Others	21
Average B value (Å <sup>2</sup> )	
Protein	99.13
main chain	97.56
side chain	100.80
Others	82.02
R.m.s. deviations	
Bonds (Å)	0.011
Angle (°)	1.473
Ramachandran plot statistics (%)	
Most favorable	91.7
Additionally allowed	8.1
Generously allowed	0.3
Disallowed	0.0

Five crystals were used for data collection. Values in parentheses are for the highest resolution shell.



Extended Data Table 2 | Summary of disease-related sequence variations of GLUT1

Variants	Disease	Variants	Disease
N34I <sup>36</sup>	DS2	R223P <sup>41</sup>	EIG12
N34S <sup>37</sup>	DS1	R223W <sup>40</sup>	DS1
N34Y <sup>40</sup>	DS1	R232C <sup>44</sup>	EIG12
S66F <sup>34</sup>	DS1	K256E <sup>34</sup>	DS1
G91D <sup>41</sup>	DS1	A275T <sup>38</sup>	DS2
R92W <sup>39</sup>	DS2	Δ282-285 <sup>38</sup>	DS2
R93W <sup>40</sup>	DS2	Y292YY <sup>57</sup>	DS1
S95I <sup>41</sup>	DS2	S294P <sup>43</sup>	DS2
M96V <sup>40</sup>	DS1	T295M <sup>40</sup>	DS1
R126C <sup>58</sup>	DS1	V303L <sup>40</sup>	DS1
R126H <sup>58</sup>	DS1	T310I <sup>7</sup>	DS1
R126L <sup>34</sup>	DS1	G314S <sup>41</sup>	DS2
G130S <sup>40</sup>	DS1	N317T <sup>59</sup>	DS2
E146K <sup>58</sup>	DS1	S324L <sup>41</sup>	DS2
R153C <sup>58</sup>	DS1	E329Q <sup>40</sup>	DS1
R153H <sup>40</sup>	DS2	R333Q <sup>40</sup>	DS1
A155V <sup>40</sup>	DS1	R333W <sup>34</sup>	DS1
V165I <sup>42</sup>	DS2	G382D <sup>40</sup>	DS1
Δ169 <sup>37</sup>	DS1	A405D <sup>40</sup>	DS1
R212C <sup>40</sup>	DS1	R468W <sup>60</sup>	DS1
R212H <sup>40</sup>	DS1	P485L <sup>40</sup>	DS1

DS1/2 stands for GLUT1 deficiency syndrome 1 or 2. EIG12 is short for 'epilepsy, idiopathic generalized 12'. The superscript numbers following each mutation are reference numbers (refs 7, 34, 36–44, 57–60).

# Dynamically important magnetic fields near accreting supermassive black holes

M. Zamaninasab<sup>1</sup>, E. Clausen-Brown<sup>1</sup>, T. Savolainen<sup>1</sup> & A. Tchekhovskoy<sup>2,3</sup>

Accreting supermassive black holes at the centres of active galaxies often produce ‘jets’—collimated bipolar outflows of relativistic particles<sup>1</sup>. Magnetic fields probably play a critical role in jet formation<sup>2,3</sup> and in accretion disk physics<sup>4</sup>. A dynamically important magnetic field was recently found near the Galactic Centre black hole<sup>5</sup>. If this is common and if the field continues to near the black hole event horizon, disk structures will be affected, invalidating assumptions made in standard models<sup>3,6,7</sup>. Here we report that jet magnetic field and accretion disk luminosity are tightly correlated over seven orders of magnitude for a sample of 76 radio-loud active galaxies. We conclude that the jet-launching regions of these radio-loud galaxies are threaded by dynamically important fields, which will affect the disk properties. These fields obstruct gas infall, compress the accretion disk vertically, slow down the disk rotation by carrying away its angular momentum in an outflow<sup>3</sup> and determine the directionality of jets<sup>8</sup>.

General relativistic magnetohydrodynamic simulations find that if the accretion disk is threaded with enough poloidal magnetic flux, the flux will be transported inwards and accumulate in the central region of the disk until the ram pressure of the accreting gas is balanced by the magnetic pressure<sup>3,7</sup>. (Poloidal refers to a vector pointing along the radial and axial directions in a cylindrical coordinate system, with the axial direction aligned with the disk angular momentum axis.) The poloidal flux threading the black hole,  $\Phi_{\text{BH}}$ , then naturally reaches a saturation or equilibrium value of  $\sim 50 \left( \dot{M} r_g^2 c \right)^{1/2}$ , where  $\dot{M}$  is the mass accretion rate,  $r_g = GM/c^2$  the black hole gravitational radius,  $G$  the gravitational constant,  $M$  the black hole mass, and  $c$  the speed of light. The magnetic field then dominates the plasma dynamics of the inner disk and modifies the disk structure such that it becomes a ‘magnetically arrested disk’<sup>3,6,7</sup>. Simulations of magnetically arrested disks also find that if the black hole’s

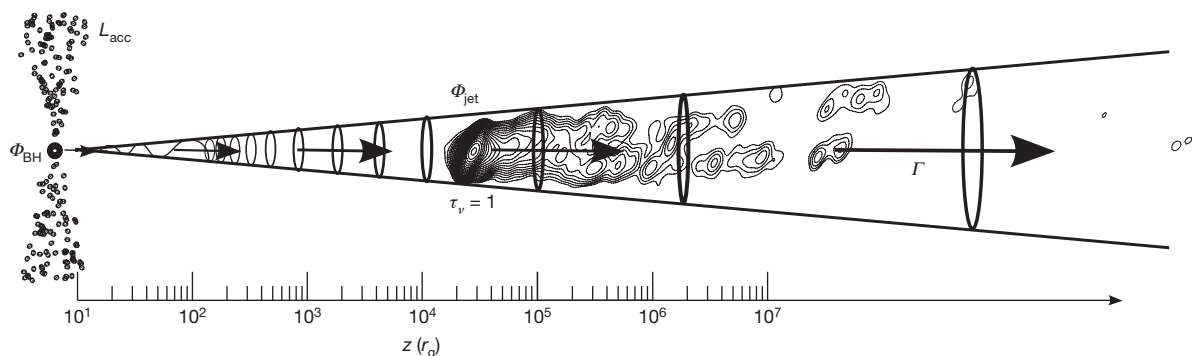
dimensionless spin parameter is  $a_* \gtrsim 0.5$  ( $a_* = 0$  refers to a non-rotating black hole and  $a_* = 1$  a maximally spinning black hole), then highly relativistic bipolar jets powered by the black hole’s rotational energy via the Blandford–Znajek effect<sup>2</sup> are ejected<sup>9,10</sup> with a total power of  $\sim \dot{M} c^2$ . Among active galactic nuclei (AGN), only about 10% contain powerful radio-emitting jets (that is, are radio-loud AGN)<sup>11</sup>.

Although we cannot directly measure  $\Phi_{\text{BH}}$ , we can observationally infer the poloidal magnetic flux threading parsec-scale jets,  $\Phi_{\text{jet}}$ , which by the flux freezing approximation is the same as  $\Phi_{\text{BH}}$  for jets produced via the Blandford–Znajek effect (Fig. 1). This inference can be performed via high-angular-resolution radio observations of the so-called core-shift effect<sup>12</sup> (see Methods), which gives a measure of the jet’s co-moving azimuthal magnetic field,  $B'_\phi = B_\phi / \Gamma$  (where primes refer to jet co-moving frame quantities, and  $\Gamma$  is jet bulk Lorentz factor). According to standard Blandford–Znajek jet theory,  $B_\phi / B_p \propto a_* R_j / r_H$ , where  $R_j$  is jet cylindrical radius,  $B_p$  is the poloidal magnetic field and  $r_H = r_g \left( 1 + (1 - a_*^2)^{1/2} \right)$  is the black hole event horizon radius. Therefore,  $\Phi_{\text{jet}} \approx R_j^2 B_p \propto M R_j B_p$ , and a more detailed derivation at one parsec downstream of the black hole yields:

$$\Phi_{\text{jet}} = 1.2 \times 10^{34} f(a_*) \Gamma \theta_j \left[ \frac{M}{10^9 M_\odot} \right] \left[ \frac{B'_{1\text{pc}}}{1 \text{ G}} \right] \quad (1)$$

Here  $f(a_*) = a_*^{-1} r_H / r_g$ ,  $B'_{1\text{pc}}$  is the jet’s co-moving frame magnetic field measured by the core-shift effect,  $\theta_j$  is the jet opening angle and  $\Phi_{\text{jet}}$  is in units of  $\text{G cm}^2$  (for a full derivation of equation (1), see Methods).

The amount of magnetic flux predicted to thread a black hole in the magnetically arrested disk model can be calculated for a given source if its mass and accretion disk luminosity are known. This predicted flux



**Figure 1 | Diagram of an AGN jet.** At left, the large black dot represents the black hole, with the disk (seen side-on) extending up and down;  $z$  (horizontal axis) represents the distance above the disk, in units of  $r_g$ . The radio core is the bright feature at the upstream end of the parsec-scale jet in the embedded radio image and it corresponds to the place in the jet where the optical depth  $\tau_\nu = 1$  for an observing frequency,  $\nu$ . The frequency-dependent positional shift of this core provides an estimate for the magnetic field at one parsec,  $B'_{1\text{pc}}$ , which we use to derive the poloidal magnetic flux threading parsec-scale jets,  $\Phi_{\text{jet}}$ . This,

by the flux freezing approximation, should be the same as the magnetic flux threading the black hole,  $\Phi_{\text{BH}}$ . We compare  $\Phi_{\text{jet}}$  to the predicted amount of magnetic flux corresponding to a magnetically arrested disk,  $50 \left( \dot{M} r_g^2 c \right)^{1/2} \propto L_{\text{acc}}^{1/2} M$ , where  $L_{\text{acc}}$  is the accretion disk luminosity.  $\Gamma$  is the jet bulk Lorentz factor and characterizes the jet velocity. This figure makes use of data from the MOJAVE database (<http://www.physics.purdue.edu/astro/mojave/>).

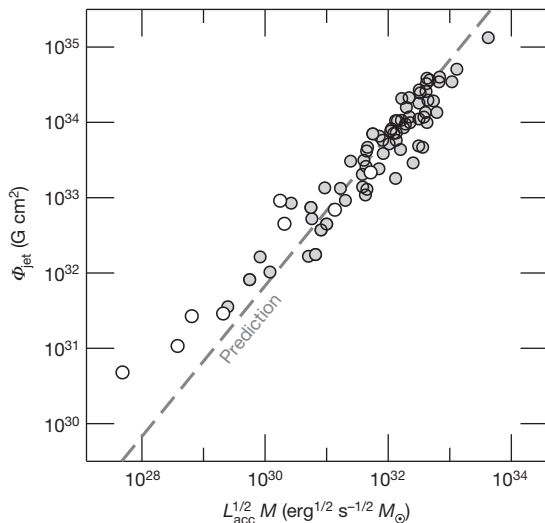
<sup>1</sup>Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany. <sup>2</sup>Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA. <sup>3</sup>Department of Astronomy and Theoretical Astrophysics Center, University of California, Berkeley, California 94720-3411, USA.

scales as  $50(\dot{M}r_g^2c)^{1/2} \propto L_{\text{acc}}^{1/2}M$ , where  $L_{\text{acc}}$  is the accretion disk luminosity. The disk luminosity enters this relation via the radiative efficiency  $\eta$ , where  $L_{\text{acc}} = \eta\dot{M}c^2$ ; in this work we use  $\eta = 0.4$  (see Methods for details).

We test the magnetically arrested disk model by plotting the values of  $\Phi_{\text{jet}}$  versus  $L_{\text{acc}}^{1/2}M$  for a sample of radio-loud AGN with measured core-shifts and accretion disk luminosities. If all sources contain magnetically arrested disks, then the data in this plot will display two qualities: (1) there will be a positive correlation between  $\Phi_{\text{jet}}$  and  $L_{\text{acc}}^{1/2}M$ , and (2) the data points will be scattered around the theoretical curve defined by the relation  $\Phi_{\text{jet}} \approx 50(\dot{M}r_g^2c)^{1/2}$ .

We use measured core-shifts from a sample of 191 radio-loud AGN<sup>13</sup> and from individual nearby sources<sup>12,14–18</sup> (see Extended Data Tables 1 and 2). We have also searched the literature for available estimates of black hole mass for the above-mentioned radio sources and found reliable estimates for 76 of them (see Methods). The masses are estimated assuming that the dynamics of the broad line region around the central black hole are governed by the black hole's gravity. Well-established empirical correlations are then used to relate the observed luminosity of the optical lines to the size of the broad line region, while line widths are used to estimate the gas velocity in this region. From the size and velocity estimates the virial mass for the central body can be calculated<sup>19–21</sup>. We have also used reported<sup>19–22</sup> broad line luminosities as a proxy for the accretion disk luminosities,  $L_{\text{acc}}$ .

We calculate  $\Phi_{\text{jet}}$  and  $L_{\text{acc}}^{1/2}M$  for each source in our sample, assuming  $a_* = 1.0$ , and plot our results in Fig. 2. A partial Kendall correlation test between  $\Phi_{\text{jet}}$  and  $L_{\text{acc}}^{1/2}M$ , considering the common dependence of these two quantities on redshift and mass, confirms a positive correlation (over  $3\sigma$  significance), which is not an artefact caused by either distance- or mass-driven selection biases (see Methods and Extended Data Tables 3 and 4 for more details). The scatter in this correlation may be caused by deviations from our assumptions that  $a_* = 1$  and  $\eta = 0.4$ , as well as by observational errors. For example, most black hole spins are predicted<sup>23</sup> to fall in the range  $0.3 \lesssim a_* \lesssim 1$ , and  $\eta$  may be different for each source, especially for low-luminosity AGN in our sample which are expected<sup>24</sup> to have  $\eta \ll 0.4$ . Given that our sample is made up of diverse sources with different accretion histories, this correlation might seem surprising. However, magnetically arrested disk theory nicely explains this because it predicts a saturation value of magnetic flux that is mostly independent of initial conditions<sup>10</sup>.



**Figure 2 | Measured magnetic flux of the jet,  $\Phi_{\text{jet}}$ , versus  $L_{\text{acc}}^{1/2}M$ .** Here we assume that  $\Gamma\theta_j = 1$ ; we also assume an accretion radiative efficiency of  $\eta = 0.4$  for our sample of 76 sources. The dashed line shows the theoretical prediction based on the magnetically arrested disk model. Filled and open circles represent blazars and radio galaxies, respectively (see Methods for details).

We fit the equation  $\Phi_{\text{jet}} = \phi_{\text{BH}}(\dot{M}r_g^2c)^{1/2}$  to the data by varying the dimensionless free parameter  $\phi_{\text{BH}}$ , which can be interpreted as a dimensionless magnetic flux and is predicted to be  $\phi_{\text{BH}} \approx 50$  for magnetically arrested disks. We find a best fit,  $\phi_{\text{BH}} = (52 \pm 5)\Gamma\theta_j$ , where the error is the  $1\sigma$  confidence interval. The best fit value for  $\phi_{\text{BH}}$  is expressed in terms of  $\Gamma\theta_j$  because this value is not well known. AGN jet launching models<sup>25</sup> typically find  $\Gamma\theta_j \approx 1$ , implying that most of the sources are close to the magnetically arrested disk prediction,  $\phi_{\text{BH}} \approx 50$ , and their magnetic fields are dynamically important. However, we note that in the less luminous regions of jets that are well downstream of the bright radio core where our magnetic field measurements are taken, the typical value of  $\Gamma\theta_j$  is<sup>26</sup> 0.1–0.2, which would give  $\phi_{\text{BH}} \approx 10$  for most sources. Even if this low value of  $\Gamma\theta_j$  holds in the radio cores, it would still imply that a fraction of the sources, including the M87 galaxy, have dynamically important magnetic fields near their central black holes. Because our magnetic field measurements use emission located closer to the jet launching regions, we assume  $\Gamma\theta_j \approx 1$  as predicted by jet launching models, and therefore  $\phi_{\text{BH}} \approx 50$  in agreement with magnetically arrested disk predictions. Thus, our data confirm the magnetically arrested disk model because it displays, as discussed above, a positive correlation between  $\Phi_{\text{jet}}$  and  $L_{\text{acc}}^{1/2}M$ , and the data points are scattered around the theoretical curve defined by  $\Phi_{\text{jet}} \approx 50(\dot{M}r_g^2c)^{1/2}$ .

Our data provide direct observational evidence that the inner accretion disks of radio-loud AGN contain strong, dynamically important magnetic fields regulated by the mass accretion rate. As most models of black hole accretion disks rely on the assumption that the magnetic pressure in the disk body is much less than the plasma pressure, our findings imply that these models may require significant changes. In particular, attempts to model the silhouette of the central black hole in the M87 galaxy and the spectral energy distributions of X-ray binaries hosting strong radio jets may be in need of significant revision. Models of the Galactic Centre accretion disk may also need to be revised, as a dynamically important magnetic field has been reported<sup>5</sup> within a distance of  $\sim 3 \times 10^7 r_g$  from the central black hole.

The quantitative agreement between  $\Phi_{\text{jet}}$  and  $50(\dot{M}r_g^2c)^{1/2}$  demonstrates that our current theoretical models of magnetically arrested disks capture the most important processes in the accretion flow responsible for jet formation. This also implies that most, if not all, radio-loud objects contain dynamically important magnetic flux near their central black holes and that it is the magnetic fields twisted by the rotation of these black holes that power their jets (that is, the Blandford–Znajek mechanism<sup>2</sup>). These twisted large-scale magnetic fields transfer energy (in the form of Poynting flux) from the rotating black holes out to parsec-scale distances, where their strength can be estimated by the core-shift effect. Our results are consistent with the proposal that radio-loud AGN consist of those black hole systems whose environment/accretion history is conducive to the formation of magnetically arrested disks, whereas radio-quiet AGN (that is, AGN without powerful jets) have failed to form magnetically arrested disks<sup>27</sup>. The idea that radio-quiet AGN are failed magnetically arrested disks is bolstered by the few studies of black hole spin made for radio-quiet AGN<sup>28</sup>, which find close to maximally spinning black holes, implying that the Blandford–Znajek power (which presumably controls radio-loudness) must be low due to low magnetic flux threading their black holes. The importance of black hole feedback in part depends on jet power, which is typically assumed<sup>29,30</sup> to be  $\sim 0.1\dot{M}c^2$ . However, our findings imply that much higher jet powers of  $\sim \dot{M}c^2$  are common<sup>3</sup>, hence suggesting that jets play a more important role in the process of AGN feedback than typically assumed.

## METHODS SUMMARY

Our sample consists of two types of source: (1) blazars (68 sources), which are AGN with jets directed almost at Earth such that their emission is significantly boosted by relativistic effects, and (2) nearby radio galaxies (8 sources), which are AGN with jets that are typically more closely aligned with the plane of the sky.



To obtain  $\Phi_{\text{jet}}$  for each source, we use published measurements of the core-shift, which is the angular distance that the radio core of each source shifts on the sky when observed at different frequencies using very long baseline interferometry (VLBI). We calculate the jet co-moving frame magnetic field strength one parsec downstream of the black hole,  $B'_{\text{pc}}$ , by using standard theoretical assumptions about the radio core's emission and absorption mechanism (the synchrotron process), the ratio of magnetic energy to particle energy, and the flux freezing approximation<sup>12</sup>. The results of analytical and numerical models of jets<sup>2,3</sup> are then used to calculate  $\Phi_{\text{jet}}$  from  $B'_{\text{pc}}$ .

For each source, we calculate  $\dot{M}$  from the  $L_{\text{acc}}$  value derived from that source's line luminosities (H $\beta$  or Mg II or O III), assuming that all the sources in our sample have the same radiative efficiency of  $\eta = 0.4$ . Although there are theoretical arguments that magnetically arrested disks could have slightly higher or lower  $\eta$ , our results are not particularly sensitive to the assumed value of  $\eta$  because the value of  $\phi_{\text{BH}}$ , which we set out to measure, is proportional to  $\eta^{1/2}$ .

The correlation we find between  $\Phi_{\text{jet}}$  and  $L_{\text{acc}}^{1/2} \dot{M}$  (in excess of  $3\sigma$ ) accounts for all of our data including a number of upper limits. We quantify the scatter in our correlation by finding the two theoretical curves defined by  $\phi_{\text{BH}}$  that encompass the middle 68% of our data, which gives  $\phi_{\text{BH}} = 24$  and  $\phi_{\text{BH}} = 93$ .

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper

Received 13 January; accepted 14 April 2014.

- Meier, D. L. *Black Hole Astrophysics: The Engine Paradigm* (Springer, 2012).
- Blandford, R. D. & Znajek, R. L. Electromagnetic extraction of energy from Kerr black holes. *Mon. Not. R. Astron. Soc.* **179**, 433–456 (1977).
- Tchekhovskoy, A., Narayan, R. & McKinney, J. C. Efficient generation of jets from magnetically arrested accretion on a rapidly spinning black hole. *Mon. Not. R. Astron. Soc.* **418**, L79–L83 (2011).
- Balbus, S. A. & Hawley, J. F. A powerful local shear instability in weakly magnetized disks. I. Linear analysis. *Astrophys. J.* **376**, 214–222 (1991); A powerful local shear instability in weakly magnetized disks. II. Nonlinear evolution. *Astrophys. J.* **376**, 223–233 (1991).
- Eatough, R. P. *et al.* A strong magnetic field around the supermassive black hole at the centre of the Galaxy. *Nature* **501**, 391–394 (2013).
- Narayan, R., Igumenshchev, I. V. & Abramowicz, M. A. Magnetically arrested disk: an energetically efficient accretion flow. *Publ. Astron. Soc. Jpn* **55**, L69–L72 (2003).
- McKinney, J. C., Tchekhovskoy, A. & Blandford, R. D. General relativistic magnetohydrodynamic simulations of magnetically choked accretion flows around black holes. *Mon. Not. R. Astron. Soc.* **423**, 3083–3117 (2012).
- McKinney, J. C., Tchekhovskoy, A. & Blandford, R. D. Alignment of magnetized accretion disks and relativistic jets with spinning black holes. *Science* **339**, 49–52 (2013).
- Tchekhovskoy, A., McKinney, J. C. & Narayan, R. General relativistic modeling of magnetized jets from accreting black holes. *J. Phys. Conf. Ser.* **372**, 012040 (2012).
- Tchekhovskoy, A. & McKinney, J. C. Prograde and retrograde black holes: whose jet is more powerful? *Mon. Not. R. Astron. Soc.* **423**, L55–L59 (2012).
- Sikora, M., Stawarz, L. & Lasota, J.-P. Radio loudness of active galactic nuclei: observational facts and theoretical implications. *Astrophys. J.* **658**, 815–828 (2007).
- Lobanov, A. P. Ultracompact jets in active galactic nuclei. *Astron. Astrophys.* **330**, 79–89 (1998).
- Pushkarev, A. B. *et al.* MOJAVE: Monitoring of jets in active galactic nuclei with VLBA experiments. IX. Nuclear opacity. *Astron. Astrophys.* **545**, A113 (2012).
- Kovalev, Y. Y., Lobanov, A. P., Pushkarev, A. B. & Zensus, J. A. Opacity in compact extragalactic radio sources and its effect on astrophysical and astrometric studies. *Astron. Astrophys.* **483**, 759–768 (2008).
- Müller, C. *et al.* Dual-frequency VLBI study of Centaurus A on sub-parsec scales. The highest-resolution view of an extragalactic jet. *Astron. Astrophys.* **530**, L11 (2011).
- Kadler, M., Ros, E., Lobanov, A. P., Falcke, H. & Zensus, J. A. The twin-jet system in NGC 1052: VLBI-scrutiny of the obscuring torus. *Astron. Astrophys.* **426**, 481–493 (2004).
- Hada, K. *et al.* An origin of the radio jet in M87 at the location of the central black hole. *Nature* **477**, 185–187 (2011).
- Martí-Vidal, I. *et al.* Detection of jet precession in the active nucleus of M81. *Astron. Astrophys.* **533**, A111 (2011).
- Kaspi, S. *et al.* Reverberation measurements for 17 quasars and the size-mass-luminosity relations in active galactic nuclei. *Astrophys. J.* **533**, 631–649 (2000).
- Woo, J.-H. & Urry, C. M. Active galactic nucleus black hole masses and bolometric luminosities. *Astrophys. J.* **579**, 530–544 (2002).
- Liu, Y., Jiang, D. R. & Gu, M. F. The jet power, radio loudness, and black hole mass in radio-loud active galactic nuclei. *Astrophys. J.* **637**, 669–681 (2006).
- Punsly, B. & Zhang, S. Calibrating emission lines as quasar bolometers. *Mon. Not. R. Astron. Soc.* **412**, L123–L127 (2011).
- Volonteri, M., Sikora, M., Lasota, J.-P. & Merloni, A. The evolution of active galactic nuclei and their spins. *Astrophys. J.* **775**, 94 (2013).
- Ho, L. C. Nuclear activity in nearby galaxies. *Annu. Rev. Astron. Astrophys.* **46**, 475–539 (2008).
- Komissarov, S. S., Vlahakis, N., Königl, A. & Barkov, M. V. Magnetic acceleration of ultrarelativistic jets in gamma-ray burst sources. *Mon. Not. R. Astron. Soc.* **394**, 1182–1212 (2009).
- Clausen-Brown, E., Savolainen, T., Pushkarev, A. B., Kovalev, Y. Y. & Zensus, J. A. Causal connection in parsec-scale relativistic jets: results from the MOJAVE VLBI survey. *Astron. Astrophys.* **558**, A144 (2013).
- Sikora, M. & Begelman, M. C. Magnetic flux paradigm for radio loudness of active galactic nuclei. *Astrophys. J.* **764**, L24–L28 (2013).
- Fabian, A. C. *et al.* Broad line emission from iron K- and L-shell transitions in the active galaxy 1H0707–495. *Nature* **459**, 540–542 (2009).
- McKinney, J. C. Total and jet Blandford-Znajek power in the presence of an accretion disk. *Astrophys. J.* **630**, L5–L8 (2005).
- Hawley, J. F. & Krolik, J. H. Magnetically driven jets in the Kerr metric. *Astrophys. J.* **641**, 103–116 (2006).

**Acknowledgements** We thank A. Lobanov, M. Sikora and J. McKinney for discussions and N. Sabha for comments on the manuscript. M.Z. was supported by the German Deutsche Forschungsgemeinschaft, DFG, via grant SFB 956 project A2. A.T. was supported by a Princeton Center for Theoretical Science fellowship and by NASA through the Einstein fellowship program, grant PF3-140115.

**Author Contributions** M.Z. proposed the experiment, compiled the data and performed most of the analysis. E.C.-B. wrote most of the main text and contributed to the theoretical analysis and implications of the work. T.S. contributed to the analysis and discussion of the data, and A.T. contributed to the theoretical analysis and implications of the work. All authors contributed ideas, discussed the results and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.Z. ([m.zamaninasab@gmail.com](mailto:m.zamaninasab@gmail.com)).

## METHODS

**Data.** Our sample of radio-loud AGN can be subdivided into two classes of objects: blazars and nearby radio galaxies. Extended Data Table 1 lists the properties of 68 blazars in our sample, including several well studied ones (3C 454.3, 3C 279 and BL Lac)<sup>15</sup>. We searched the literature and found optical line observations that we used to calculate the mass of the corresponding black holes and the bolometric luminosities of their accretion disks.

Extended Data Table 2 lists the same properties as Extended Data Table 1, but for nearby radio galaxies that have core-shifts and accretion disk luminosities reported in the literature (we found only eight such cases). Core-shift measures are taken from different individual experiments: NGC 1052<sup>16</sup>, 3C 120<sup>14</sup>, M 81<sup>18</sup>, W Comae<sup>14</sup>, M 87<sup>17</sup>, Cen A<sup>15</sup>, 3C 390.3<sup>14</sup> and Cyg A<sup>12</sup>. Owing to their proximity, six of these objects have black hole masses measured either via stellar velocity dispersion/nuclear gas dynamics modelling (NGC 1052, M 81, M 87, Cen A, Cyg A) or reverberation mapping (3C 120).

**Calculating the jet poloidal magnetic flux from the core-shift measure.** According to the standard model of relativistic jets<sup>31</sup>, the observed position of the radio core (a bright, synchrotron self-absorbed emission feature at the upstream end of the jet) is a function of the observing frequency. This frequency-dependent shift in the location of the core can be used to estimate the magnetic field strength and electron number density<sup>12,32</sup>. The method is based on several simple and reasonable assumptions: the flow has a conical shape with a small half-opening angle,  $\theta_j$ , and constant  $\Gamma$  (no significant acceleration or deceleration). Furthermore, the magnetic field strength and electron number density are assumed to decrease with increasing distance from the central engine following power-law dependencies:  $B'(z) = B'_{1pc}(z/1pc)^{-\alpha_b}$ ,  $n'_e(z) = n'_{e,1pc}(z/1pc)^{-\alpha_n}$  where  $z$  is the distance from the black hole (see Fig. 1). Here  $B'_{1pc}$  and  $n'_{e,1pc}$  represent the jet frame magnetic field strength and electron number density at distance of one parsec away from the apex of the jet. These assumptions lead to a power-law behaviour<sup>12</sup> for the amount of core-shift as a function of the observing frequency,  $\propto \nu^{-1/k_r}$ , where  $k_r$  is the core-shift index. We calculate the core-shift using<sup>12</sup>:

$$\Omega_{rv} = 4.85 \times 10^{-9} \left[ \frac{\Delta r_{core, v_1 v_2} D_L}{(1+z_*)^2} \right] \left[ \frac{v_1^{1/k_r} v_2^{1/k_r}}{v_2^{1/k_r} - v_1^{1/k_r}} \right] \text{ pc GHz} \quad (2)$$

where  $\Delta r_{core, v_1 v_2}$ ,  $D_L$  and  $z_*$  are the observed shift (in milliarcsec) of the core position between frequencies  $v_1$  and  $v_2$  (in GHz), the luminosity distance, and redshift of the source respectively (note we use  $z_*$  for redshift to distinguish it from the cylindrical coordinate  $z$ ).

If the magnetic field decays with distance as  $z^{-1}$  as expected for an azimuthally dominated magnetic field, and the magnetic to non-thermal particle energy ratio remains constant, then it is expected that  $k_r = 1$ . Numerous observations have indeed revealed a  $k_r = 1$  behaviour of the core-shift, thus supporting our assumption of a azimuthally dominated magnetic field in the radio core region<sup>14,33,34</sup>. Assuming equipartition between magnetic and particle energy at the position of the radio core, then with the above assumptions for the scaling of the magnetic field and relativistic electron density with distance, one can estimate the jet frame magnetic field strength at 1 pc from the black hole:

$$B'_{1pc} \approx 0.025 \left[ \frac{\sigma_{rel} \Omega_{rv}^3 (1+z_*)^2}{\theta_j^2 \delta^2 \sin^2 \theta} \right]^{0.25} \quad (3)$$

where  $\delta$  and  $\theta$  are the Doppler factor and the viewing angle of the jet, respectively. It is assumed that the source has a spectral index of  $\alpha = -0.5$ . Here  $\sigma_{rel}$  is the ratio of the energy density in the magnetic field to the energy density in the non-thermal relativistic particles responsible for jet emission. We assume equipartition such that  $\sigma_{rel} = 1$ , though if this assumption is incorrect it would not strongly affect our results because the jet magnetic flux goes as  $\Phi_{jet} \propto \sigma_{rel}^{1/4}$ .

For the estimate of the magnetic field we assume that (when a good estimate of the viewing angle is not available) jets of this sample of blazars are viewed close to their critical angle<sup>13</sup>,  $\theta \approx \Gamma^{-1}$ . We compared a sub-sample of 32 sources with known Doppler factors<sup>35</sup> and viewing and intrinsic opening angles<sup>26,36,37</sup> and found this is in fact a reasonable assumption for this sample. This yields<sup>13</sup>:

$$B'_{1pc} \approx 0.042 \Omega_{rv}^{3/4} (1+z_*)^{1/2} (1+\beta_{app}^2)^{1/8} \quad (4)$$

where  $\beta_{app}$  is the characteristic apparent speed of a jet's components.

We assume the mean jet frame magnetic field measured in the radio core is predominantly azimuthal, as expected in magnetically powered relativistic jets<sup>38-40</sup>. As discussed above, this assumption is supported by the findings from core-shift studies that  $k_r = 1$ , as well as other VLBI studies<sup>33</sup> of jets that are consistent with a  $B \propto z^{-1}$  scaling, which is the expected scaling for a azimuthally dominated magnetic field in a conical jet. Theoretically, in Blandford-Znajek jets the azimuthal to poloidal field

ratio is  $B_\phi/B_p \approx R_j/R_{LC} \approx a_* R_j / [(4-8)r_g]$ , which clearly is much greater than unity on VLBI scales<sup>2</sup>. This ratio was calculated for a rapidly spinning black hole ( $a_* \approx 1$ ), and  $R_{LC} = c/\Omega_F$  is the radius of the light cylinder, where  $\Omega_F$  is the angular frequency of magnetic field lines that remains constant along the jet (it is probably<sup>3,41,42</sup> between 1/4 and 1/2 of the black hole angular frequency  $\Omega_H = ca_*/(2r_H)$ ).

Let us consider a representative field line in the bulk of the jet, where most of the energy flows out and the observer frame magnetic field is strongest. In an idealized case of a highly magnetized jet, the poloidal magnetic field is approximately uniform over the cross-section of the jet, and we have two familiar relationships: the observer frame azimuthal field scales as  $B_\phi \propto R_j^{-1}$ , and the poloidal field scales as  $B_p \propto R_j^{-2}$ , where  $R_j$  is the jet cylindrical radius<sup>43</sup>. Realistic jets are, however, probably not very highly magnetized: the process of jet acceleration requires conversion of Poynting flux to kinetic flux. In this conversion process,  $B_p$  becomes non-uniform across the jet<sup>25,44</sup>. Because of this,  $B_p$  decreases faster than  $R_j^{-2}$  and  $B_\phi$  decreases faster than  $R_j^{-1}$  in the jet acceleration region. The poloidal magnetic field strength  $B_p$  along the field line is related<sup>44</sup> to the total enclosed poloidal magnetic flux in the jet by  $\Phi_{jet} = k\pi R_j^2 B_p$ , where  $k = (1 - \Gamma/\mu)^{-1} \gtrsim 1$  is the correction factor and  $\Gamma$  is the Lorentz factor of the field line. Here  $\mu$  is the Lorentz factor that the field line would have if all of the Poynting flux were converted into kinetic energy flux. If we now eliminate  $B_p$  in favour of  $B_\phi$  using the relationship in the previous paragraph, we obtain a more general form for equation (1):

$$\Phi_{jet} = \frac{2\pi(\Gamma\theta_j)r_H z B'_\phi}{\ell a_* (1 - \Gamma/\mu)} \quad (5)$$

where  $z$  is the distance down the jet,  $B'_\phi = B_\phi/\Gamma$  is the azimuthal magnetic field in the jet frame, and  $\ell$  is defined such that  $\Omega_F = \ell\Omega_H$ . In order to reproduce equation (1), we assume  $z = 1$  pc,  $\ell = 1/2$ ,  $\Gamma = \mu/2$ , and  $B'_\phi \approx B'_{1pc}$  because the jet co-moving frame magnetic field probed by core-shift measurements on parsec-scales,  $B'_{1pc}$ , is likely to be azimuthally dominated. The assumption that  $\Gamma = \mu/2$  is justified by different analytical and numerical<sup>25,44</sup> studies of jet acceleration, which have found that at the end of the acceleration region roughly 50% of a jet's Poynting flux is converted into kinetic energy flux. We convert between jet cylindrical radius  $R_j$  and jet downstream distance  $z$  by using a conical jet approximation such that  $R_j \approx \theta_j z$ .

**Predicted magnetic flux in a magnetically arrested disk.** Under the assumption that the width of the optical broad lines emitted from the clouds surrounding black holes is mainly regulated by the gravitational potential of the black holes (assuming that these ionized clouds follow Keplerian orbits) and using a well-established empirical relation between the size of the broad line region and optical continuum luminosity, one can derive an order of magnitude estimate for black hole masses (for different line diagnostics)<sup>19,45-50</sup>:

$$M \approx 4.82 \times \left[ \frac{\lambda L_{5100\text{\AA}}}{10^{44} \text{ erg s}^{-1}} \right]^{0.69} \left[ \frac{\text{FWHM}_{H\beta}}{\text{km s}^{-1}} \right]^2 M_\odot \quad (6)$$

$$M \approx 3.37 \times \left[ \frac{\lambda L_{3000\text{\AA}}}{10^{44} \text{ erg s}^{-1}} \right]^{0.47} \left[ \frac{\text{FWHM}_{MgII}}{\text{km s}^{-1}} \right]^2 M_\odot \quad (7)$$

$$M \approx 33.1 \times \left[ \frac{\lambda L_{1350\text{\AA}}}{10^{44} \text{ erg s}^{-1}} \right]^{0.46} \left[ \frac{\text{FWHM}_{CIV}}{\text{km s}^{-1}} \right]^2 M_\odot \quad (8)$$

Here  $\lambda$  is the observed monochromatic continuum luminosity at wavelength  $\lambda$  and  $\text{FWHM}_{line}$  is the observed full-width at half-maximum of the line in question. In radio-loud AGN, the optical emission can be contaminated by the non-thermal synchrotron emission of their jets, so we estimate the optical continuum on the basis of the broad emission line luminosities<sup>21</sup>.

Furthermore, we have used broad (and in two cases narrow) line luminosities as a surrogate for the thermal (non-beamed) emission from the accretion disk. The scaling relations between  $L_{acc}$  and the line luminosity,  $L_{line}$ , for different lines are<sup>19,22,51,52</sup>:

$$\log_{10} L_{acc} = (12.32 \pm 0.32) + (0.78 \pm 0.01) \log_{10} L_{H\beta} \quad (9)$$

$$\log_{10} L_{acc} = (16.76 \pm 0.26) + (0.68 \pm 0.01) \log_{10} L_{MgII} \quad (10)$$

$$\log_{10} L_{acc} = (26.50 \pm 0.32) + (0.46 \pm 0.01) \log_{10} L_{OIII} \quad (11)$$

We may now calculate the predicted magnetically arrested disk magnetic flux for a source with known  $M$  and  $L_{acc}$ :

$$50 \left( \dot{M} r_g^2 c \right)^{1/2} = 2.4 \times 10^{34} \left[ \frac{\eta}{0.4} \right]^{-1/2} \left[ \frac{M}{10^9 M_\odot} \right] \left[ \frac{L_{acc}}{1.26 \times 10^{47} \text{ erg s}^{-1}} \right]^{1/2} \text{ G cm}^2 \quad (12)$$

To calculate  $\dot{M}$  from  $L_{acc}$ , we assume that all of the sources in our sample have the same radiative efficiency  $\eta$ . If the gravitational binding energy of accreting gas is

radiated away by the time it reaches the black hole's innermost stable orbit, then for a maximally spinning black hole<sup>53</sup>  $\eta \approx 0.42$ . Magnetically arrested disks may behave differently from standard disks, with  $\eta \approx 0.5$  for non-rotating black holes and potentially higher  $\eta$  for rotating black holes<sup>6</sup>. The radiative efficiency could be even higher if inverse Compton scattering by the hot, magnetized disk corona reprocesses disk emission. On the other hand, if a fraction of liberated binding energy is carried away mechanically (for example, by magnetized winds) instead of by radiation, the resulting  $\eta$  values could be lower (since the efficiency of magnetized winds in the magnetically arrested disk model is<sup>9</sup>  $\eta \leq 0.2$ , we expect this to be at most a modest correction). Note that since the value of  $\phi_{\text{BH}}$ , which we set out to measure, is proportional to  $\eta^{1/2}$ , our results are not particularly sensitive to the assumed value of  $\eta$ . In this work, we adopt a representative value,  $\eta = 0.4$ .

**Linear regression and partial correlation analysis.** In order to probe for possible correlations between observed values (for example, accretion luminosity and magnetic field strength), one should take into account their possible common dependencies on other parameters (for example, distance, mass and relativistic boosting). We have done this by using a generalized version of the partial Kendall's  $\tau_K$  correlation test<sup>54</sup> that is suitable for our sample (which includes a number of upper limits). We used the publicly available FORTRAN code<sup>55</sup> which has been widely used for probing correlations in astronomical data sets. Extended Data Table 3 summarizes the results of our partial correlation analysis.

Furthermore, we have double-checked our results with the ppcorr package<sup>56</sup> which is available as a public library for  $\mathcal{R}$ <sup>57</sup>. The ppcorr package calculates the Kendall  $\tau_K$  correlation between different observed values while controlling for dependencies on more than one commonly dependent variable. However, ppcorr does not take upper and lower limits (so-called censored data points) into account properly. Extended Data Table 4 summarizes the results of a test using ppcorr for a correlation between  $\Phi_{\text{jet}}$  and  $L_{\text{acc}}^{1/2} M$  while simultaneously controlling for common dependencies on mass, redshift and apparent velocity.

The scatter in the observed correlation can be quantified in terms of the free parameter defining the theoretically predicted curves,  $\phi_{\text{BH}}$ . If we search for the two theoretical curves that encompass the middle 68% of the data points (that is, 16% of the data lie above the upper line and 16% of the data lie below the lower line), then we find the curves defined by  $\phi_{\text{BH}} = 24$  and  $\phi_{\text{BH}} = 93$  satisfy this criterion (recall  $\phi_{\text{BH}} \approx 50$  is the value predicted by magnetically arrested disk theory). Note that when we state the results of the fit,  $\phi_{\text{BH}} = (52 \pm 5)$ , the error represents the confidence interval of the best fit value, and does not refer to the scatter discussed above. The best fit value and its confidence interval are obtained using the MPFIT package<sup>58,59</sup>.

We also performed linear regression between different observed values (for example,  $\Phi_{\text{jet}}$  and  $L_{\text{acc}}^{1/2} M$ ) using a Bayesian method that accurately accounts for the censored data points<sup>60</sup>. For this purpose we have used the LINMIX\_ERR routine which is publicly available as a part of the astrolib package<sup>61</sup>. We have used the Metropolis-Hastings algorithm available in the code to obtain the posterior distribution. The linear regression results in a best-fit model  $\log \Phi_{\text{jet}} = (8.52 \pm 1.06) + (0.79 \pm 0.03) \times \log L_{\text{acc}}^{1/2} M$  (1 $\sigma$  confidence).

31. Blandford, R. D. & Königl, A. Relativistic jets as compact radio sources. *Astrophys. J.* **232**, 34–48 (1979).
32. Hirotani, K. Kinetic luminosity and composition of active galactic nuclei jets. *Astrophys. J.* **619**, 73–85 (2005).
33. O'Sullivan, S. P. & Gabuzda, D. C. Magnetic field strength and spectral distribution of six parsec-scale active galactic nuclei jets. *Mon. Not. R. Astron. Soc.* **400**, 26–42 (2009).
34. Sokolovsky, K. V., Kovalev, Y. Y., Pushkarev, A. B. & Lobanov, A. P. A VLBA survey of the core shift effect in AGN jets. I. Evidence of dominating synchrotron opacity. *Astron. Astrophys.* **532**, A38 (2011).
35. Hovatta, T., Valtaoja, E., Tornikoski, M. & Lähteenmäki, A. Doppler factors, Lorentz factors and viewing angles for quasars, BL Lacertae objects and radio galaxies. *Astron. Astrophys.* **494**, 527–537 (2009).
36. Pushkarev, A. B., Kovalev, Y. Y., Lister, M. L. & Savolainen, T. Jet opening angles and gamma-ray brightness of AGN. *Astron. Astrophys.* **507**, L33–L36 (2009).
37. Savolainen, T. *et al.* Relativistic beaming and gamma-ray brightness of blazars. *Astron. Astrophys.* **512**, A24 (2010).
38. Gabuzda, D. C., Vitrichshak, V. M., Mahmud, M. & O'Sullivan, S. P. Radio circular polarization produced in helical magnetic fields in eight active galactic nuclei. *Mon. Not. R. Astron. Soc.* **384**, 1003–1014 (2008).
39. Pudritz, R. E., Hardcastle, M. J. & Gabuzda, D. C. Magnetic fields in astrophysical jets: from launch to termination. *Space Sci. Rev.* **169**, 27–72 (2012).
40. Zamaninasab, M. *et al.* Evidence for a large-scale helical magnetic field in the quasar 3C 454.3. *Mon. Not. R. Astron. Soc.* **436**, 3341–3356 (2013).
41. McKinney, J. C. General relativistic magnetohydrodynamic simulations of the jet formation and large-scale propagation from black hole accretion systems. *Mon. Not. R. Astron. Soc.* **368**, 1561–1582 (2006).
42. Tchekhovskoy, A., Narayan, R. & McKinney, J. C. Black hole spin and the radio loud/quiet dichotomy of active galactic nuclei. *Astrophys. J.* **711**, 50–63 (2010).
43. Tchekhovskoy, A., McKinney, J. C. & Narayan, R. Simulations of ultrarelativistic magnetodynamic jets from gamma-ray burst engines. *Mon. Not. R. Astron. Soc.* **388**, 551–572 (2008).
44. Tchekhovskoy, A., McKinney, J. C. & Narayan, R. Efficiency of magnetic to kinetic energy conversion in a monopole magnetosphere. *Astrophys. J.* **699**, 1789–1808 (2009).
45. Wandel, A., Peterson, B. M. & Malkan, M. A. Central masses and broad-line region sizes of active galactic nuclei. I. Comparing the photoionization and reverberation techniques. *Astrophys. J.* **526**, 579–591 (1999).
46. McLure, R. J. & Jarvis, M. J. Measuring the black hole masses of high-redshift quasars. *Mon. Not. R. Astron. Soc.* **337**, 109–116 (2002).
47. Kaspi, S. *et al.* The relationship between luminosity and broad-line region size in active galactic nuclei. *Astrophys. J.* **629**, 61–71 (2005).
48. Greene, J. E. & Ho, L. C. Estimating black hole masses in active galaxies using the H $\alpha$  emission line. *Astrophys. J.* **630**, 122–129 (2005).
49. Shen, Y. *et al.* A catalog of quasar properties from Sloan Digital Sky Survey Data Release 7. *Astrophys. J.* **194**, 45 (2011).
50. Shaw, M. S. *et al.* Spectroscopy of broad-line blazars from 1LAC. *Astrophys. J.* **748**, 49 (2012).
51. Celotti, A., Padovani, P. & Ghisellini, G. Jets and accretion processes in active galactic nuclei: further clues. *Mon. Not. R. Astron. Soc.* **286**, 415–424 (1997).
52. Gu, M., Cao, X. & Jiang, D. R. The bulk kinetic power of radio jets in active galactic nuclei. *Mon. Not. R. Astron. Soc.* **396**, 984–996 (2009).
53. Novikov, I. D. & Thorne, K. S. in *Black Holes* (eds Dewitt, C. & Dewitt, B. S.) 343–450 (Gordon and Breach, Paris, 1973).
54. Akritas, M. G. & Siebert, J. A test for partial correlation with censored astronomical data. *Mon. Not. R. Astron. Soc.* **278**, 919–924 (1996).
55. [http://astro.psu.edu/statcodes/cens\\_tau.f](http://astro.psu.edu/statcodes/cens_tau.f).
56. <http://cran.r-project.org/web/packages/ppcor/index.html>.
57. <http://www.r-project.org>.
58. Markwardt, C. B. in *Astronomical Data Analysis Software and Systems XVIII* (eds Bohlender, D. A., Durand, D. & Dowler, P.) 251–254 (Astron. Soc. Pacif. Conf. Ser. Vol. 411, Astronomical Society of Pacific, 2009).
59. <http://purl.org/net/mpfit>.
60. Kelly, B. C. Some aspects of measurement error in linear regression of astronomical data. *Astrophys. J.* **665**, 1489–1506 (2007).
61. <http://idlastro.gsfc.nasa.gov>.
62. Torrealba, J. *et al.* Optical spectroscopic atlas of the MOJAVE/2cm AGN sample. *Rev. Mex. Astron. Astrofis.* **48**, 9–40 (2012).
63. Palma, N. I. *et al.* Multiwavelength observations of the gamma-ray blazar PKS 0528+134 in quiescence. *Astrophys. J.* **735**, 60 (2011).
64. Di Matteo, T., Allen, S. W., Fabian, A. C., Wilson, A. S. & Young, A. J. Accretion onto the supermassive black hole in M87. *Astrophys. J.* **582**, 133–140 (2003).
65. Evans, D. A. *et al.* Chandra and XMM-Newton observations of the nucleus of Centaurus A. *Astrophys. J.* **612**, 786–796 (2004).
66. Gebhardt, K. *et al.* The black hole mass in M87 from Gemini/NIFS adaptive optics observations. *Astrophys. J.* **729**, 119 (2011).
67. Ghisellini, G. *et al.* General physical properties of bright Fermi blazars. *Mon. Not. R. Astron. Soc.* **402**, 497–518 (2010).
68. González-Martín, O., Masegosa, J., Márquez, I. & Guainazzi, M. Fitting LINER nuclei within the active galactic nucleus family: a matter of obscuration? *Astrophys. J.* **704**, 1570–1585 (2009).
69. King, A. L. *et al.* What is on tap? The role of spin in compact objects and relativistic jets. *Astrophys. J.* **771**, 84 (2013).
70. Markoff, S. *et al.* Results from an extensive simultaneous broadband campaign on the underluminous active nucleus M81\*: further evidence for mass-scaling accretion in black holes. *Astrophys. J.* **681**, 905–924 (2008).
71. McNamara, B. R., Rohanizadegan, M. & Nulsen, P. E. J. Are radio active galactic nuclei powered by accretion or black hole spin? *Astrophys. J.* **727**, 39 (2011).
72. Neumayer, N. *et al.* The central parsecs of Centaurus A: high-excitation gas, a molecular disk, and the mass of the black hole. *Astrophys. J.* **671**, 1329–1344 (2007).
73. Kino, M. & Kawakatu, N. Estimate of the total kinetic power and age of an extragalactic jet by its cocoon dynamics: the case of Cygnus A. *Mon. Not. R. Astron. Soc.* **364**, 659–664 (2005).



Extended Data Table 1 | Blazar sample

Object (1)	Type (2)	$z_*$ (3)	$\beta_{\text{app}}$ (c) (4)	$\Omega_{r\nu}$ (pc GHz) (5)	$B'_{1\text{pc}}$ (G) (6)	$\text{Log}_{10} M (M_{\odot})$ (7)	$\text{Log}_{10} L_{\text{acc}}$ (erg s $^{-1}$ ) (8)	Reference (9)
0133+476	Q	0.86	12.98	13.69	0.77	8.55	46.11	Tor12
0212+735	Q	2.37	7.64	21.24	1.27	9.96	47.33	Tor12
0234+285	Q	1.21	12.26	35.82	1.71	9.22	46.80	Shw12
0333+321	Q	1.26	12.76	41.51	1.95	9.25	47.17	Liu06
0336+019	Q	0.85	22.36	14.41	0.92	8.89	46.33	Liu06
0403+132	Q	0.57	19.69	33.30	1.54	9.08	46.52	Liu06
0420+014	Q	0.91	7.36	35.94	1.41	8.41	46.48	Liu06
0454+844	B	0.11	0.14	10.91	0.27	7.42	44.66	Tor12
0528+134	Q	2.07	19.20	22.71	1.60	9.03	47.23	Pal11
0605+085	Q	0.87	19.79	13.24	0.84	8.87	46.44	Shw12
0607+157	Q	0.32	3.93	21.22	0.68	7.32	45.20	Liu06
0736+017	Q	0.19	14.32	<2.94	<0.20	7.86	45.69	Liu06
0738+313	Q	0.63	10.76	16.85	0.81	9.57	46.93	Liu06
0748+126	Q	0.89	18.37	13.49	0.84	9.06	46.88	Tor12
0804+499	Q	1.44	1.83	11.15	0.48	9.39	46.80	Liu06
0827+243	Q	0.94	22.01	19.64	1.18	8.89	46.46	Tor12
0836+710	Q	2.22	25.38	25.72	1.93	9.36	47.51	Liu06
0859+140	Q	1.34	16.47	30.96	1.70	9.14	46.73	Tor12
0906+015	Q	1.02	20.68	29.45	1.61	8.55	46.61	Liu06
0917+624	Q	1.45	15.57	16.99	1.09	8.93	46.57	Shw12
0923+392	Q	0.69	4.29	<6.64	<0.33	9.09	46.94	Liu06
0945+408	Q	1.25	18.60	17.02	1.10	8.99	46.46	Tor12
0953+254	Q	0.71	11.52	<6.71	<0.42	8.70	46.30	Liu06
1015+359	Q	1.23	12.46	15.55	0.92	8.69	46.65	Tor12
1038+064	Q	1.26	11.87	21.97	1.19	9.12	46.76	Tor12
1127+145	Q	1.18	14.18	13.24	0.84	9.30	46.87	Tor12
1156+295	Q	0.73	24.73	20.11	1.17	8.54	46.25	Liu06
1219+044	Q	0.96	2.35	24.16	0.81	8.89	46.48	Tor12
1222+216	Q	0.43	21.10	17.03	0.90	8.87	46.34	Tor12
1253+055	Q	0.54	20.57	<5.88	<0.42	8.28	46.05	Liu06
1302+102	Q	0.28	5.41	20.34	0.70	8.51	46.26	Liu06
1308+326	Q	1.00	27.17	13.61	0.96	8.72	46.38	Tor12
1334+127	Q	0.54	10.26	31.08	1.23	7.98	45.98	Liu06
1458+718	Q	0.90	7.04	9.46	0.51	9.23	46.80	Liu06
1502+106	Q	1.84	14.77	8.50	0.69	8.74	46.92	Liu06
1508+055	Q	1.19	18.64	26.81	1.52	9.32	46.71	Tor12
1510+089	Q	0.36	20.14	13.50	0.73	8.20	46.05	Liu06
1546+027	Q	0.41	12.08	<5.09	<0.32	8.47	46.32	Liu06
1606+106	Q	1.23	18.91	10.97	0.79	8.97	46.56	Tor12
1611+343	Q	1.40	14.11	9.07	0.66	9.19	46.78	Tor12
1633+382	Q	1.81	29.45	21.21	1.62	9.12	46.80	Tor12
1637+574	Q	0.75	10.61	13.51	0.71	9.22	46.77	Liu06
1641+399	Q	0.59	19.27	23.85	1.20	9.27	46.69	Liu06
1642+690	Q	0.75	16.65	<6.85	<0.48	6.81	45.17	Tor12
1655+077	Q	0.62	14.45	7.45	0.47	7.28	45.60	Liu06
1726+455	Q	0.72	1.82	<6.73	<0.28	9.18	46.63	Tor12
1749+701	B	0.77	6.03	27.03	1.04	8.77	45.96	Tor12
1803+784	B	0.68	8.97	<6.58	<0.39	7.92	45.98	Liu06
1807+698	B	0.05	0.10	3.81	0.12	8.51	44.98	Tor12
1823+568	B	0.66	20.85	11.79	0.74	7.94	45.61	Tor12
1828+487	Q	0.69	13.65	12.24	0.69	8.41	46.34	Tor12
1849+670	Q	0.66	30.63	6.48	0.52	8.81	46.22	Tor12
1901+319	Q	0.63	2.67	29.03	0.87	7.72	46.08	Tor12
1928+738	Q	0.30	8.43	12.31	0.54	8.35	46.47	Liu06
2121+053	Q	1.94	13.29	22.61	1.43	8.78	47.15	Liu06
2128+123	Q	0.50	6.94	26.41	0.98	9.02	46.65	Liu06
2134+004	Q	1.93	5.94	26.23	1.31	9.36	46.93	Tor12
2145+067	Q	0.99	2.52	<7.48	<0.34	8.87	47.07	Liu06
2155+152	Q	0.67	18.11	43.23	1.89	7.59	45.67	Woo02
2200+420	B	0.07	10.57	1.21	0.09	8.23	45.18	Tor12
2201+315	Q	0.29	7.87	27.04	0.95	8.94	46.68	Liu06
2209+236	Q	1.12	3.43	7.69	0.39	8.46	46.39	Shw12
2227+088	Q	1.56	8.14	29.60	1.44	8.98	46.58	Tor12
2230+114	Q	1.04	15.41	46.48	2.12	8.93	46.58	Tor12
2243+123	Q	0.63	5.49	20.10	0.78	8.81	46.64	Tor12
2251+158	Q	0.86	14.19	22.00	0.32	8.69	46.86	Liu06
2345+167	Q	0.58	13.45	18.44	0.90	8.47	45.84	Liu06
2351+456	Q	1.99	18.01	25.97	1.72	9.29	46.68	Tor12

(1), IAU name (B1950); (2), classification of the source (Q, quasar; B, BL Lac object; RG, radio galaxy); (3), redshift; (4), apparent velocity of the jet; (5), core-shift measure; (6), magnetic field strength one parsec away from the jet base; (7), mass of the black hole; (8), accretion disk luminosity; (9), reference for the line used for estimating mass and accretion luminosity. References: Liu06, ref. 21; Tor12, ref. 62; Shw12, ref. 50; Pal11, ref. 63; Woo02, ref. 20.

Extended Data Table 2 | Radio galaxy sample

Object (1)	Type (2)	$z_*$ (3)	$\beta_{\text{app}}$ (c) (4)	$\Omega_{r\nu}$ (pc GHz) (5)	$B'_{\text{ipc}}$ (G) (6)	$\text{Log}_{10} M$ ( $M_{\odot}$ ) (7)	$\text{Log}_{10} L_{\text{acc}}$ (erg s $^{-1}$ ) (8)	Reference (9)
0238+084 (NGC 1052)	RG	0.0050	0.23	0.226	0.015	8.19	41.24	Woo02/Gon09
0430+052 (3C 120)	RG	0.033	4.66	1.85	0.11	7.36	43.92	King13
0951+693 (M 81)	RG	0.0001	1.45	0.031	0.0066	7.80	39.77	Mar08
1219+285 (W Coma)	RG	0.102	4.05	1.39	0.08	8.69	43.25	Ghe10
1228+126 (M 87)	RG	0.004	0.24	0.132	0.012	9.82	40.85	Geb11/DiM03
1322+427 (Cen A)	RG	0.0018	0.50	0.155	0.017	7.74	41.68	Neu09/Eva04
1845+797 (3C 390.3)	RG	0.057	2.29	1.65	0.089	8.83	44.61	Tor12
1957+405 (Cyg A)	RG	0.0562	0.20	2.10	0.07	9.43	44.57	McN11/Kino05

Similar to Extended Data Table 1. References: DiM03, ref. 64; Eva04, ref. 65; Geb11, ref. 66; Ghe10, ref. 67; Gon09, ref. 68; King13, ref. 69; Mar08, ref. 70; McN11, ref. 71; Neu09, ref. 72; Kino05, ref. 73; Tor12, ref. 62; Woo02, ref. 20.

**Extended Data Table 3 | Partial correlation analysis results based on cens\_tau algorithm**

Variables			Correlation		Significance of rejecting the null hypothesis
$X_1$	$X_2$	$X_3$	$\tau_K$	$P_{\text{null}}$	
All sources:					
Log $L_{\text{acc}}^{1/2} M$	Log $\Phi_{\text{jet}}$	$z_*$	0.62	$< 10^{-13}$	$> 5 \sigma$
Log $L_{\text{acc}}^{1/2} M$	Log $\Phi_{\text{jet}}$	Log $M$	0.46	$1.1 \times 10^{-3}$	$3.3 \sigma$
Log $L_{\text{acc}}^{1/2} M$	Log $\Phi_{\text{jet}}$	$\beta_{\text{app}}$	0.72	$< 10^{-14}$	$> 5 \sigma$
Excluding LLAGN:					
Log $L_{\text{acc}}^{1/2} M$	Log $\Phi_{\text{jet}}$	$z_*$	0.61	$< 10^{-13}$	$> 5 \sigma$
Log $L_{\text{acc}}^{1/2} M$	Log $\Phi_{\text{jet}}$	Log $M$	0.21	$1.2 \times 10^{-3}$	$3.2 \sigma$
Log $L_{\text{acc}}^{1/2} M$	Log $\Phi_{\text{jet}}$	$\beta_{\text{app}}$	0.70	$< 10^{-14}$	$> 5 \sigma$

$X_1$  and  $X_2$  are variables that have a possible mutual dependency on variable  $X_3$ .  $\tau_K$  and  $P_{\text{null}}$  are the Kendall's partial correlation coefficient and probability for the null hypothesis of no correlation between  $X_1$  and  $X_2$ , respectively. The last column shows the corresponding significance by which the null hypothesis could be rejected.



Extended Data Table 4 | Results of partial correlation analysis based on ppcorr algorithm

	$X_1$	$X_2$	Variables			Correlation		Significance of rejecting the null hypothesis
			$X_3$	$X_4$	$X_5$	$\tau_K$	$P_{\text{null}}$	
Excluding upper limits:	$\text{Log } L_{\text{acc}}^{1/2} M$	$\text{Log } \Phi_{\text{jet}}$	$z_*$	$\text{Log } M$	$\beta_{\text{app}}$	0.52	$1.9 \times 10^{-9}$	$> 5 \sigma$
Excluding upper limits and LLAGN:	$\text{Log } L_{\text{acc}}^{1/2} M$	$\text{Log } \Phi_{\text{jet}}$	$z_*$	$\text{Log } M$	$\beta_{\text{app}}$	0.32	$5.2 \times 10^{-3}$	$3.0 \sigma$

$X_1$  and  $X_2$  are variables that have a possible mutual dependency on variables  $X_3$ ,  $X_4$  and  $X_5$ .  $\tau_K$  and  $P_{\text{null}}$  are the same as in Extended Data Table 3.

# Palladium-catalysed C–H activation of aliphatic amines to give strained nitrogen heterocycles

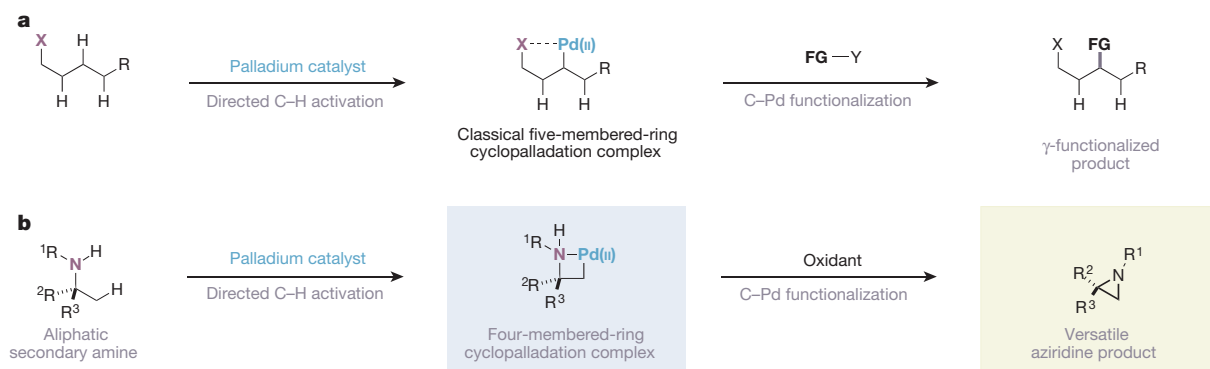
Andrew McNally<sup>1</sup>, Benjamin Haffemayer<sup>1</sup>, Beatrice S. L. Collins<sup>1</sup> & Matthew J. Gaunt<sup>1</sup>

The development of new chemical transformations based on catalytic functionalization of unactivated C–H bonds has the potential to simplify the synthesis of complex molecules dramatically. Transition metal catalysis has emerged as a powerful tool with which to convert these unreactive bonds into carbon–carbon and carbon–heteroatom bonds<sup>1–6</sup>, but the selective transformation of aliphatic C–H bonds is still a challenge. The most successful approaches involve a ‘directing group’, which positions the metal catalyst near a particular C–H bond, so that the C–H functionalization step occurs via cyclometallation<sup>7</sup>. Most directed aliphatic C–H activation processes proceed through a five-membered-ring cyclometallated intermediate<sup>8–10</sup>. Considering the number of new reactions that have arisen from such intermediates, it seems likely that identification of distinct cyclometallation pathways would lead to the development of other useful chemical transformations<sup>11</sup>. Here we report a palladium-catalysed C–H bond activation mode that proceeds through a four-membered-ring cyclopalladation pathway. The chemistry described here leads to the selective transformation of a methyl group that is adjacent to an unprotected secondary amine into a synthetically versatile nitrogen heterocycle. The scope of this previously unknown bond disconnection is highlighted through the development of C–H amination and carbonylation processes, leading to the synthesis of aziridines and  $\beta$ -lactams (respectively), and is suggestive of a generic C–H functionalization platform that could simplify the synthesis of aliphatic secondary amines, a class of small molecules that are particularly important features of many pharmaceutical agents.

Several classes of directing group can participate in directed C–H activation of aliphatic systems and include functionalities such as carbonyl derivatives, aromatic nitrogen heterocycles and hydroxyl motifs<sup>12–18</sup>. Palladium salts have been particularly successful at effecting cyclometallation with these functional groups; coordination of the palladium centre to the Lewis basic heteroatom within the directing motifs is assumed to lower the entropic and enthalpic costs of the C–H bond cleavage and ring

closure, and accounts for why kinetically favoured five-membered-ring intermediates are preferred over other cyclopalladation pathways. Reaction of the resulting carbon–palladium bond with an external reagent leads to an overall chemical transformation that sees a carbon–hydrogen bond converted into a synthetically versatile functional group. As a result, this type of cyclopalladation event has led to a number of useful catalytic C–H bond functionalization processes that have expanded the toolbox of reactions available to synthetic chemists (Fig. 1a). Despite their success, one limitation of catalytic cyclopalladation is that the exquisite selectivity displayed through five-membered-ring intermediates means that the generic strategic bond disconnection enabled by directed C–H activation is often restricted to a single position with respect to the controlling functional group. As part of an overarching goal to develop new activation modes for catalytic C–H activation, we questioned whether cyclopalladation could function in systems that do not possess C–H bonds in the positions that would facilitate classical, kinetically favoured insertion processes. Furthermore, we reasoned that C–H activation directed by common functional groups represents an important challenge to the continued advance of this field; whereas native functionalities such as carboxylic acids and hydroxyl groups have been shown to participate in catalytic cyclopalladation<sup>17,18</sup>, there are no examples of directed catalytic C–H activation with unprotected aliphatic amines<sup>19–23</sup>. Here we report the successful realization of these ideas via a new C–H activation strategy which uses an unprotected aliphatic secondary amine to direct a palladium-catalysed C–H activation event that transforms an adjacent methyl group into a synthetically versatile nitrogen heterocycle (Fig. 1b). This novel palladium-catalysed activation process proceeds through a remarkable four-membered-ring cyclometallation pathway and enables a broad scope for catalytic C–H functionalization through both Pd(II) or Pd(IV) intermediates.

At the outset of our studies, we selected an amine substrate displaying a series of C–H bonds positioned in such a way that they could not engage in classical five-membered-ring cyclopalladation (Fig. 2a). When



**Figure 1 | Palladium-catalysed directed C–H activation modes.** **a**, Classical directed C–H activation via a kinetically favoured five-membered-ring cyclopalladation complex. X, directing group; FG, new functional group. **b**, New directed C–H activation of an aliphatic secondary amine, via a

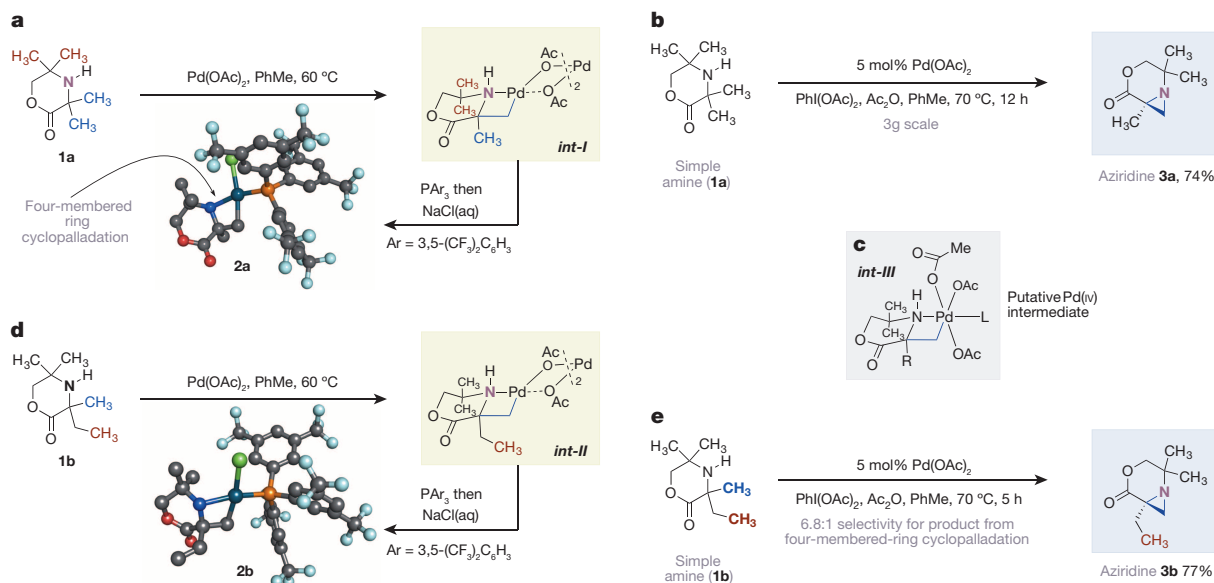
four-membered-ring cyclopalladation complex and novel reactivity in the presence of an oxidant, to form synthetically versatile strained nitrogen heterocycles, such as aziridines. R<sup>1–3</sup>, general aliphatic substituents.

<sup>1</sup>Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

the aliphatic secondary amine **1a** was treated with a stoichiometric amount of palladium(II) acetate, we observed a new trinuclear organopalladium complex (**int-I**) where C–H activation had taken place at one of the unactivated methyl groups proximal to the amine and formed a four-membered-ring intermediate (as determined by single-crystal X-ray diffraction of a phosphine derivative, **2a**). To the best of our knowledge, this is the first example of an amine-directed four-membered-ring cyclopalladation event<sup>24,25</sup>. We next probed the reactivity of these strained cyclopalladation complexes; we questioned whether mild chemical oxidants, such as the hypervalent iodine compound PhI(OAc)<sub>2</sub>, might allow access to C–H functionalization pathways proceeding through high-valent Pd(IV) cyclometallated intermediates<sup>26</sup>, as well as formulating the basis of a catalytic transformation. When we treated a cyclopalladation complex (**int-I** or **2a**) with PhI(OAc)<sub>2</sub> we were surprised to discover aziridine **3a**, presumably formed via C–N bond forming reductive elimination from a high oxidation state Pd(IV) intermediate (**int-III**, Fig. 2b, c). We established a catalytic transformation based on this unusual C–H functionalization event; after optimization, the Pd-catalysed C–H process requires only 5 mol% catalyst to produce aziridine **3a** on a multi-gram scale. Furthermore, the reaction is also selective between the two different types of methyl groups (red and blue in Fig. 1a), indicating a subtle stereoelectronic differentiation imparted by the nearby carbonyl group. The distinctive nature of the directed cyclopalladation pathway was further evidenced when we tested amine **1b**, where cyclopalladation can proceed through a four-membered-ring (activation on methyl group) or the conventional five-membered-ring pathway (activation on ethyl group) (Fig. 2d). Remarkably, we still observe activation through the four-membered-ring pathway forming the related strained complexes (**int-II**, **2b**)—we do not believe that such selectivity has been observed before. We also found that the amine **1b** underwent the same catalytic transformation in very good yield, forming the corresponding aziridine **3b** (Fig. 2e). Not only does this catalytic cyclopalladation pathway contrast with classical directed C–H activation, it also reveals novel chemical reactivity that underpins a conceptually distinct disconnection to synthesize strained nitrogen heterocycles directly from simple aliphatic amines displaying methyl groups adjacent to an unprotected nitrogen motif.

In assessing the scope of the catalytic reaction, a range of amines displaying a variety of substituents were found to be suitable substrates (Fig. 3). For example, simple alkyl-(methyl) substituted morpholinones undergo the C–H aziridination process in high yield (**3a–e**) and are always selective for the four-membered-ring C–H activation pathway. Substituents containing aromatic rings, protected hydroxyl groups, sensitive chloromethyl and fluoromethyl motifs, esters and protected amines all work well, indicating a broad tolerance of the process to a variety of useful functional groups (**3f–k**). The C–H aziridination reaction can be applied to methylene C–H bonds, albeit in lower yield, to form **3l**, further expanding the potential efficacy of this process. Morpholinones that display only three substituents around the secondary amine motif have so far not been found to react under the standard conditions. Substrates containing substituents other than methyl groups distal to the carbonyl motif can be accommodated; a piperidine substituted morpholinone undergoes smooth C–H aziridination to form a product (**3n**) displaying a high density of useful functionality. Interestingly, we found that amine derivatives not possessing the lactone framework also underwent catalytic C–H activation through the four-membered-ring cyclopalladation pathway in the presence of PhI(OAc)<sub>2</sub> but formed acetoxylation product (**5**) rather than an aziridine. This again suggests that the carbonyl group in the morpholinones may play a subtle controlling role, this time steering the reactivity of the putative organo-Pd(IV) intermediate. Importantly, acyclic amines also undergo directed C–H activation to form an amino alcohol derivative (**6**) is reasonable yield.

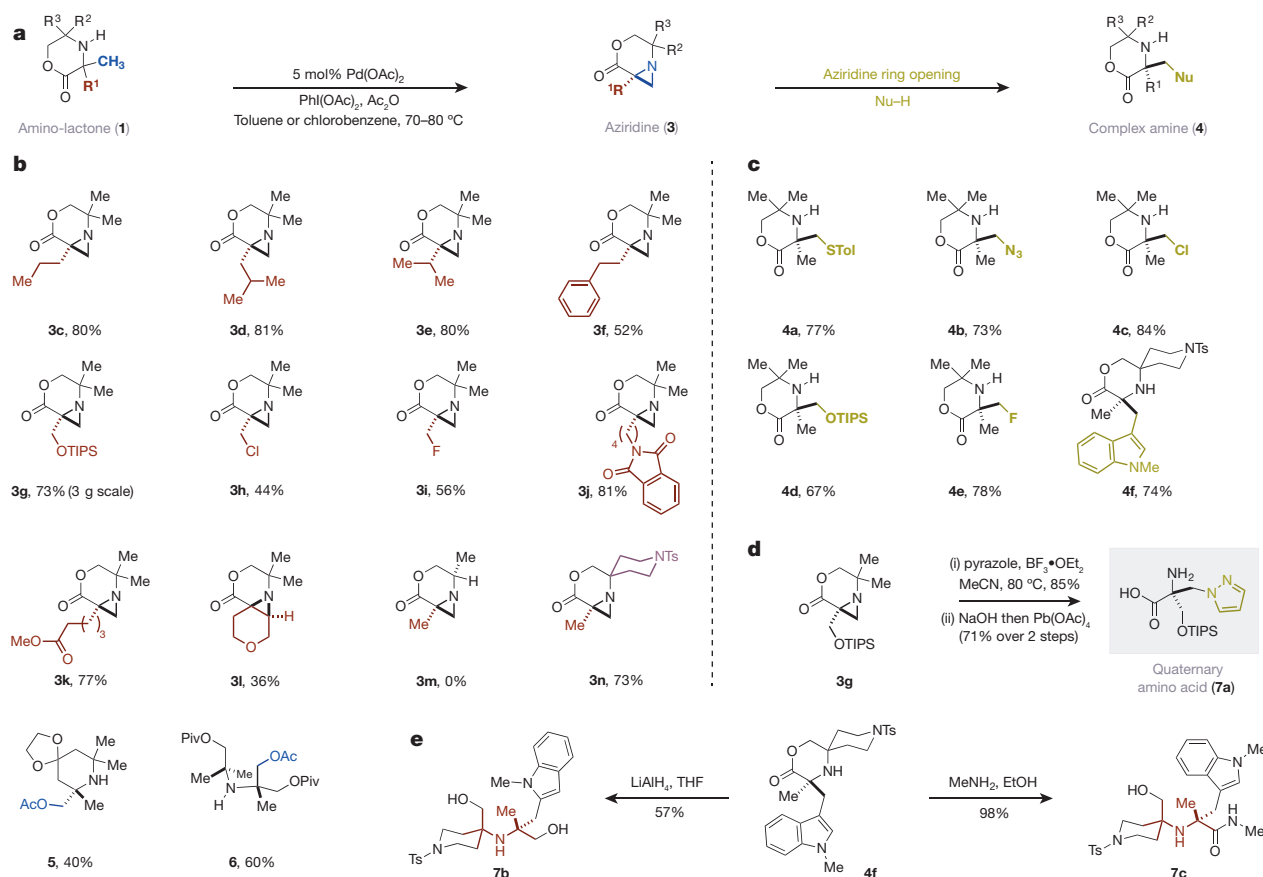
One of the most important uses of aziridines is their ring opening reactions with nucleophiles<sup>27</sup>. Despite the non-activated and hindered nature of these aziridines (**3**), we found that a range of nucleophiles can open this strained heterocycle to form complex amine products (Fig. 3c). For example, heteroatom nucleophiles such as azide, thiol, chloride and water open the aziridine in good yield under acidic conditions (**4a–d**). Notably, fluoride can also be used as a nucleophile, forming amine **4e**. The aziridine motif can also be opened with a carbon nucleophile (*N*-methylindole) to form **4f**. To further demonstrate the synthetic versatility of these products, the functionalized aziridine **3g** can also be opened with pyrazole followed by saponification-oxidative cleavage of the cyclic framework, revealing a highly functionalized quaternary amino



**Figure 2 | A new amine-directed, palladium-catalysed directed C–H activation mode.** **a**, Coordination of amine **1a** to the Pd(II) centre facilitates a cyclopalladation to give a four-membered-ring complex **int-I**. Proof of structure is obtained through X-ray diffraction of a phosphine derivative, **2a**. **b**, **e**, Palladium-catalysed C–H aziridination of secondary amines (**1a** in panel **b**, **1b** in panel **e**). Reaction to give respective aziridines **3a** and **3b** most probably proceeds through the oxidation of **int-I** to the Pd(IV) intermediate

**int-III** (**c**) before a reductive elimination to form a C–N bond. **d**, The novel cyclopalladation pathway prevails when a classical, kinetically favoured five-membered-ring cyclopalladation (**int-II**) is possible; the structure of **int-II** was determined through X-ray diffraction of a phosphine derivative, **2b**. For X-ray structures, grey is C, red is O, dark blue is N, aquamarine is Pd, orange is P, green is Cl, and light blue is F.





**Figure 3 | Scope of the palladium-catalysed C–H aziridination process.**

**a**, Transformation of aminolactone (**1**) to aziridine (**3**) and then ring opening with nucleophiles (Nu–H) gives complex amine **4**. **b**, Scope of aziridination reaction. For **5**, mass balance comprises a complex mixture of polyacetoxyated products. **c**, Scope of aziridine ring opening (see Supplementary Information

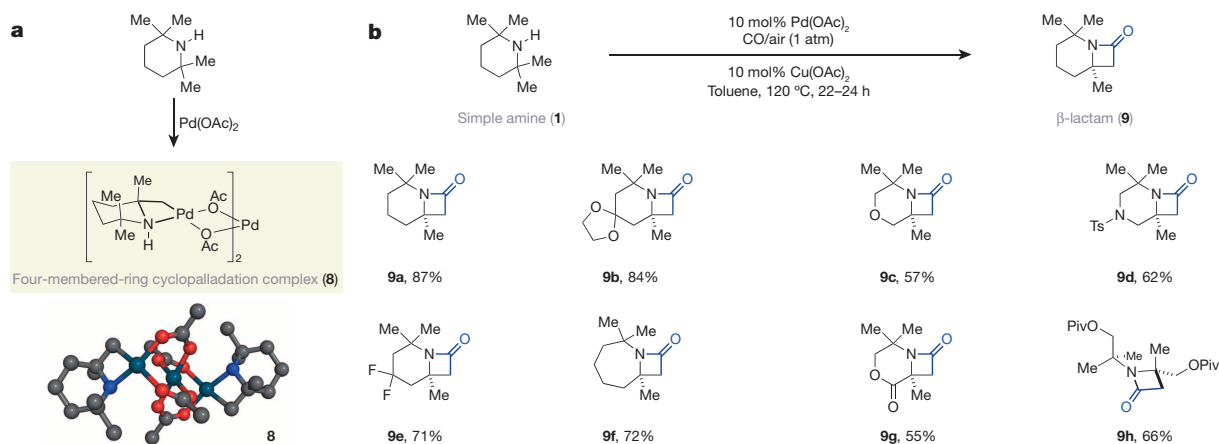
for experimental procedures). **d**, Synthesis of a functionally complex quaternary amino acid (**7a**) from **3g**. **e**, Representative transformations of the aziridine ring opening products (**4f**) into highly functionalized fully substituted secondary aliphatic amines (**7b, c**).

acid **7a**, characteristic of a class of molecule that would be difficult to make by conventional synthetic methods (Fig. 3d). Moreover, the C–H activation strategy also provides access to a previously inaccessible class of fully substituted secondary aliphatic amines. The lactone motif in morpholinones such as **4f** can be transformed into useful functionality (**7b, c**) that is representative of a class of complex fully substituted aliphatic secondary amines with unexplored biological properties; the limited number of ways through which this class of sterically hindered amine can be synthesized means they have therefore been largely ignored as pharmaceutical candidates. However, we believe that the important nature of aliphatic nitrogen-containing compounds in medical chemistry will generate significant interest in this elusive class of amines.

We next explored the reactivity of these molecules under a complementary palladium manifold. Replacing PhI(OAc)<sub>2</sub> with carbon monoxide and a suitable oxidant would lead to a potential carbonylation process through a Pd(II)–Pd(0) catalytic cycle<sup>28</sup>. We found that mixing tetramethylpiperidine (TMP) with stoichiometric Pd(OAc)<sub>2</sub> gave rise to a four-membered-ring palladacycle (**8**) that could be directly characterized by X-ray diffraction (Fig. 4a). Furthermore, when we treated this complex with carbon monoxide, a  $\beta$ -lactam (**9a**) was formed in high yield (Fig. 4b). Here again, a strained nitrogen heterocycle can be formed from a methyl group adjacent to a secondary amine, and this further underlines the synthetic versatility of this new directed C–H activation mode. Assessment of a range of reaction parameters led to an optimized catalytic process that required treatment of TMP with 10 mol% of Pd(OAc)<sub>2</sub>, and 10 mol% of Cu(OAc)<sub>2</sub> as oxidant, in toluene at 120 °C under a CO/air mixture (at atmospheric pressure for 22 h) to form  $\beta$ -lactam **9a** in 87% yield. As shown in Fig. 1b, an array of commercial

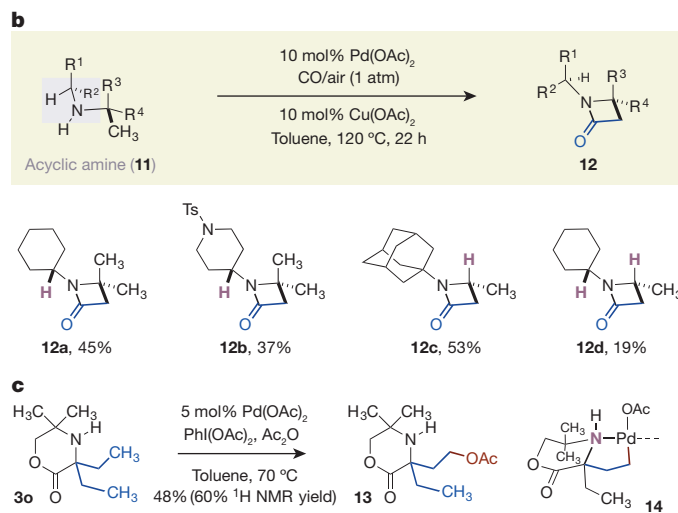
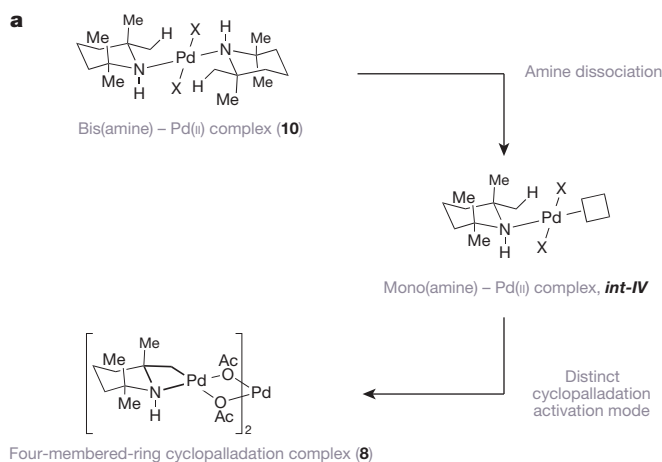
or readily available piperidine derivatives are competent substrates. Not only does this reaction produce a series of novel  $\beta$ -lactams that may themselves have interesting biological properties, but many of the products also display orthogonal and versatile functionality that can be easily transformed into more complex amines. In addition to standard piperidine derivatives (**9a, b**), a number of other heterocycles could also be readily tolerated: morpholine and piperazine scaffolds, fluorinated piperidines and a seven-membered-ring azepine all productively form the corresponding  $\beta$ -lactams and display the type of structural motifs that are frequently considered important in the design of pharmaceutical agents (**9c–f**). The morpholinone **1a** also undergoes carbonylation, but, in contrast to the C–H aziridination process, a 2:1 mixture of separable  $\beta$ -lactams, in favour of **9g**, is formed. This result is surprising given the selectivity of the cyclopalladation with this substrate and suggests that the C–H palladation step may be reversible under carbonylation conditions. Acyclic amines are also effective substrates, producing the monocyclic  $\beta$ -lactam **9h** in good yield.

Given the broad utility and application of aliphatic amines, it is surprising that they are seldom employed in C–H bond functionalization reactions, especially considering that secondary amines are excellent coordinating groups for palladium salts. When treated with palladium(II) salts, most amines rapidly form square-planar, coordinately saturated bis-amine palladium(II) complexes, but the pathway for C–H bond functionalization requires liberation of a coordination site occupied by one of these amines<sup>21</sup>. As the palladium(II) complexes formed from these amines are typically very stable, there is little driving force for the release of an amine ligand. This generally renders the bis-amine palladium species catalytically inactive unless they are subjected to thermal activation that



**Figure 4 | Palladium-catalysed C–H carbonylation of aliphatic amines.** **a**, Similarly to amine **1a**, treatment of TMP with  $\text{Pd}(\text{OAc})_2$  generates the four-membered ring cyclopalladation complex **8**, whose structure was directly determined by X-ray diffraction. **b**, Scope of the C–H carbonylation reaction.

can also lead to unwanted side reactions. The hindered secondary amines used in this study, which contain two fully substituted carbon atoms bonded to nitrogen, also form bis-amine complexes with  $\text{Pd}(\text{OAc})_2$  (**10**, Fig. 5). We propose that the steric hindrance around the  $\text{Pd}(\text{II})$  centre results in more facile dissociation. This weaker binding would therefore facilitate the release of one of the amines to create the essential vacant coordination site (*int-IV*) and enable C–H activation to take place<sup>29</sup>. We speculated that if secondary amines that do not possess two fully



Top row, reaction of **1** to give **9**; middle and bottom rows, range of products and yields. For **9g**, a 2:1 mixture of regioisomers was observed in favour of the one shown. For **9h**, the reaction was conducted with  $\text{AgOAc}$  in dioxane at  $125^\circ\text{C}$ .

substituted carbon groups directly attached to the NH motif could be accommodated (as represented by amine **11**), then the efficacy of this new C–H activation mode would be significantly expanded. Although we were conscious such amines might be prone to competitive decomposition pathways potentially via  $\beta$ -hydride elimination, we found that these less hindered amines, composed of a fully substituted carbon atom on one side of the NH motif and a tertiary carbon atom on the other, undergo the palladium-catalysed C–H carbonylation reaction to form the corresponding  $\beta$ -lactams **12a–c** (Fig. 5b). The C–H activation pathway was even viable on amines displaying two tertiary substituents (hence a C–H bond adjacent to nitrogen on both substituents), and although the yield to form **12d** is low, this initial result suggests that the generic activation pathway could be possible on many classes of secondary amine. Finally, we were also able to demonstrate that more classical five-membered-ring cyclopalladation pathways can also be initiated by this type of secondary aliphatic amine. On treatment of ethyl-substituted morpholinone **30** with  $\text{Pd}(\text{OAc})_2$  and  $\text{PhI}(\text{OAc})_2$  as oxidant, we observed a reasonable yield (55%) of the C–H acetoxylation product **13** that is presumably derived from the five-membered-ring palladacycle **14**. Based on the extensive number of transformations available to related five-membered-ring cyclopalladation pathways, this result suggests that aliphatic amines could be amenable to a range of new reactions through this well established activation mode.

Taken together, the present results represent an important initial advance that could substantially broaden this directed C–H activation methodology to further classes of secondary amine, and further reinforces the potential of a new strategic bond disconnection which converts simple methyl groups into a range of versatile strained nitrogen heterocycles. As the physical and biological properties of aliphatic amines are central to the function of many important pharmaceuticals, chemical reagents and polymeric materials, we expect that such an approach to amine synthesis will be broadly useful in the synthesis of complex functional amines.

**Figure 5 | Towards a general strategy for C–H activation of aliphatic amines.** **a**, Amines bind strongly to palladium(II) salts to form stable bis-amine complexes (**10**). Dissociation of one of the amine ligands leads to C–H activation and then (via *int-IV*) to the cyclopalladation complex **8**. The steric properties of these hindered amines are likely to be responsible for the facile dissociation; note that the  $\text{Pd}(\text{II})$  centre is projected directly into a high concentration of aliphatic C–H bonds. **b**, Top row, less hindered amines (**11**) can also undergo the directed C–H activation through the four-membered ring; bottom row, scope of products (**12a–d**). **c**, Demonstration that C–H activation of unprotected aliphatic secondary amines (such as **30**) can also proceed through five-membered-ring cyclopalladation pathways (**14**) to give compounds such as **13**.

## METHODS SUMMARY

**Pd(II)-catalysed C–H amination to aziridines.** An oven-dried round-bottom flask, equipped with stir bar, was charged with the amine (1.0 equiv.), Pd(OAc)<sub>2</sub> (0.05 equiv.) and iodobenzene diacetate (1.5–2.5 equiv.). The solvent (0.1–0.05 M) was added, followed by acetic anhydride (2.0 equiv.), and the flask capped under an air atmosphere. The flask was placed in a pre-heated oil bath at 70 °C and stirred for the given time. The reaction mixture was then cooled to room temperature, filtered through Celite, eluted with ethyl acetate, and concentrated *in vacuo*. The residue was purified by flash column chromatography to provide the pure aziridine product.

**Pd(II)-catalysed C–H carbonylation to  $\beta$ -lactams.** An oven-dried round-bottom flask, equipped with stir bar, was charged with Pd(OAc)<sub>2</sub> (0.1 equiv.), Cu(OAc)<sub>2</sub> (0.1 equiv.) and the amine (as a 0.1 M solution in toluene), and then fitted with a reflux condenser and rubber septa. A balloon of air and a balloon filled with carbon monoxide were fitted and then the flask was placed in a pre-heated oil bath at 120 °C such that the oil level matched the level of solvent in the flask and stirred for 22–24 h. The reaction mixture was then cooled to room temperature and filtered through Celite, eluting with ethyl acetate. The filtrates were washed with a 0.1 M solution of hydrochloric acid (20 ml), a saturated aqueous solution of NaHCO<sub>3</sub> (20 ml) and brine (20 ml). The organic phase was dried (MgSO<sub>4</sub>), filtered and concentrated *in vacuo*. The residue was then purified by flash column chromatography to provide the pure  $\beta$ -lactam product.

Received 23 December 2013; accepted 14 April 2014.

Published online 28 May 2014.

- Jia, C., Kitamura, T. & Fujiwara, Y. Catalytic functionalization of arenes and alkanes via C–H bond activation. *Acc. Chem. Res.* **34**, 633–639 (2001).
- Godula, K. & Sames, D. C–H bond functionalization in complex organic synthesis. *Science* **312**, 67–72 (2006).
- Davies, H. M. L. & Manning, J. R. Catalytic C–H functionalization by metal carbenoid and nitrenoid insertion. *Nature* **451**, 417–424 (2008).
- Davies, H. M. L., Du Bois, J. & Yu, J.-Q. C–H functionalization in organic synthesis. *Chem. Soc. Rev.* **40**, 1855–1856 (2011).
- Yamaguchi, J., Yamaguchi, A. D. & Itami, K. C–H bond functionalization: emerging synthetic tools for natural products and pharmaceuticals. *Angew. Chem. Int. Edn* **51**, 8960–9009 (2012).
- Mkhalid, I. A. I., Barnard, J. H., Marder, T. B., Murphy, J. M. & Hartwig, J. F. C–H activation for the construction of C–B bonds. *Chem. Rev.* **110**, 890–931 (2010).
- Dupont, J., Consorti, C. C. & Spencer, J. The potential of palladacycles: more than just precatalysts. *Chem. Rev.* **105**, 2527–2572 (2005).
- Lyons, T. W. & Sanford, M. S. Palladium-catalyzed ligand-directed C–H functionalization reactions. *Chem. Rev.* **110**, 1147–1169 (2010).
- Ackermann, L. Carboxylate-assisted transition-metal-catalyzed C–H bond functionalizations: mechanism and scope. *Chem. Rev.* **111**, 1315–1345 (2011).
- Colby, D. A., Bergman, R. G. & Ellman, J. A. Rhodium-catalyzed C–C bond formation via heteroatom directed C–H bond activation. *Chem. Rev.* **110**, 624–655 (2010).
- Leow, D., Li, G., Mei, T.-S. & Yu, J.-Q. Activation of remote meta-C–H bond assisted by an end-on template. *Nature* **486**, 518–522 (2012).
- Desai, L. V., Hull, K. L. & Sanford, M. S. Palladium-catalyzed oxygenation of unactivated sp<sup>3</sup> hybridized C–H bonds. *J. Am. Chem. Soc.* **126**, 9542–9543 (2004).
- Zaitsev, V. G., Shibashov, D. & Daugulis, O. Highly regioselective arylation of sp<sup>3</sup> C–H bonds catalyzed by palladium acetate. *J. Am. Chem. Soc.* **127**, 13154–13155 (2005).
- Wang, D.-H., Wasa, M., Giri, R. & Yu, J.-Q. Pd(II)-catalyzed cross-coupling of sp<sup>3</sup> C–H bonds with sp<sup>2</sup> and sp<sup>3</sup> boronic acids using air as the oxidant. *J. Am. Chem. Soc.* **130**, 7190–7191 (2008).
- Giri, R., Chen, X. & Yu, J.-Q. Palladium-catalyzed asymmetric iodination of unactivated C–H bonds under mild conditions. *Angew. Chem. Int. Edn* **44**, 2112–2115 (2005).
- Campeau, L.-C., Schipper, D. J. & Fagnou, K. Site-selective sp<sup>2</sup> and benzylic sp<sup>3</sup> palladium-catalyzed direct arylation. *J. Am. Chem. Soc.* **130**, 3266–3267 (2008).
- Giri, R. *et al.* Palladium-catalyzed methylation and arylation of sp<sup>2</sup> and sp<sup>3</sup> C–H bonds in simple carboxylic acids. *J. Am. Chem. Soc.* **129**, 3510–3511 (2007).
- Simmons, E. M. & Hartwig, J. F. Catalytic functionalization of unactivated primary C–H bonds directed by an alcohol. *Nature* **483**, 70–73 (2012).
- He, G., Zhao, Y., Zhang, S., Lu, C. & Chen, G. Highly efficient syntheses of azetidines, pyrrolidines, and indolines via palladium-catalyzed intramolecular amination of C(sp<sup>3</sup>)–H and C(sp<sup>2</sup>)–H bonds at the  $\gamma$  and  $\delta$  positions. *J. Am. Chem. Soc.* **134**, 3–6 (2012).
- Zhang, S.-Y. *et al.* Palladium-catalyzed picolinamide-directed alkylation of unactivated C(sp<sup>3</sup>)–H bonds with alkyl iodides. *J. Am. Chem. Soc.* **135**, 2124–2127 (2013).
- Ryabov, A. D. Mechanisms of intramolecular activation of carbon-hydrogen bonds in transition-metal complexes. *Chem. Rev.* **90**, 403–424 (1990).
- Chan, K. S. L. *et al.* Ligand-enabled cross-coupling of C(sp<sup>3</sup>)–H bonds with arylboron reagents via Pd(II)/Pd(0) catalysis. *Nature Chem.* **6**, 146–150 (2014).
- Wehn, P. M., Lee, J. & Du Bois, J. Stereochemical models for Rh-catalysed amination reactions of chiral sulfamates. *Org. Lett.* **5**, 4823–4826 (2003).
- Tan, Y., Barrios-Landeros, F. & Hartwig, J. F. Mechanistic studies on direct arylation of pyridine N-oxide: evidence for cooperative catalysis between two distinct palladium centers. *J. Am. Chem. Soc.* **134**, 3683–3686 (2012).
- Agnus, Y., Gross, M., Labarelle, M., Louis, R. & Metz, B. J. Unusual double palladium(II) cyclometallation in a tweezer-like receptor with stacked arms: synthesis, properties and X-ray structure. *J. Chem. Soc. Chem. Commun.* 939–940 (1994).
- Hickman, A. J. & Sanford, M. S. High-valent organometallic copper and palladium in catalysis. *Nature* **484**, 177–185 (2012).
- Hu, X. E. Nucleophilic ring openings of aziridines. *Tetrahedron* **60**, 2701–2743 (2004).
- Yoo, E. J., Wasa, M. & Yu, J.-Q. Pd(II)-catalyzed carbonylation of C(sp<sup>3</sup>)–H bonds: a new entry to 1,4-dicarbonyl compounds. *J. Am. Chem. Soc.* **132**, 17378–17380 (2010).
- Engle, K. M., Wu, H.-C. & Yu, J.-Q. Weak coordination as a powerful means for developing broadly useful C–H functionalization reactions. *Acc. Chem. Res.* **45**, 788–802 (2012).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We thank the EPSRC (A.M.), GSK (for a Case Award to B.H.) and the University of Cambridge (B.S.L.C.) for financial support, and the ERC and EPSRC for Fellowships (M.J.G.). Mass spectrometry data were acquired at the EPSRC UK National Mass Spectrometry Facility at Swansea University.

**Author Contributions** A.M., B.H. and B.S.L.C. discovered and developed the reactions. M.J.G. conceived, designed and directed the investigations and wrote the manuscript with revisions provided by A.M. and B.S.L.C.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.G. ([mjg32@cam.ac.uk](mailto:mjg32@cam.ac.uk)).



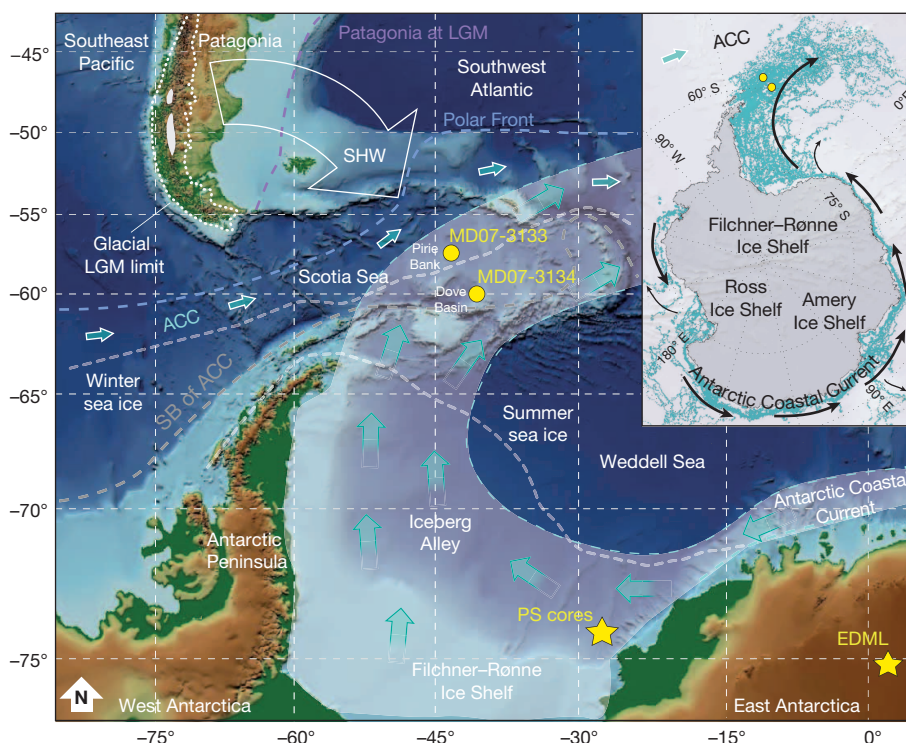
# Millennial-scale variability in Antarctic ice-sheet discharge during the last deglaciation

M. E. Weber<sup>1</sup>, P. U. Clark<sup>2</sup>, G. Kuhn<sup>3</sup>, A. Timmermann<sup>4</sup>, D. Sprenk<sup>1</sup>, R. Gladstone<sup>5</sup>, X. Zhang<sup>3</sup>, G. Lohmann<sup>3</sup>, L. Menviel<sup>6,7</sup>, M. O. Chikamoto<sup>4</sup>, T. Friedrich<sup>4</sup> & C. Ohlwein<sup>8</sup>

Our understanding of the deglacial evolution of the Antarctic Ice Sheet (AIS) following the Last Glacial Maximum (26,000–19,000 years ago)<sup>1</sup> is based largely on a few well-dated but temporally and geographically restricted terrestrial and shallow-marine sequences<sup>2–4</sup>. This sparseness limits our understanding of the dominant feedbacks between the AIS, Southern Hemisphere climate and global sea level. Marine records of iceberg-rafted debris (IBRD) provide a nearly continuous signal of ice-sheet dynamics and variability. IBRD records from the North Atlantic Ocean have been widely used to reconstruct variability in Northern Hemisphere ice sheets<sup>5</sup>, but comparable records from the Southern Ocean of the AIS are lacking because of the low resolution and large dating uncertainties in existing sediment cores. Here we present two well-dated, high-resolution IBRD records that capture a spatially integrated signal of AIS variability during the last deglaciation. We document eight events of increased iceberg flux from various parts of the AIS between 20,000 and 9,000 years ago, in marked contrast to previous scenarios which identified the main AIS retreat as occurring after meltwater pulse 1A<sup>3,6–8</sup> and continuing into the late Holocene epoch. The highest IBRD flux occurred 14,600 years

ago, providing the first direct evidence for an Antarctic contribution to meltwater pulse 1A. Climate model simulations with AIS freshwater forcing identify a positive feedback between poleward transport of Circumpolar Deep Water, subsurface warming and AIS melt, suggesting that small perturbations to the ice sheet can be substantially enhanced, providing a possible mechanism for rapid sea-level rise.

Today, the estimated total iceberg calving flux from Antarctica is about 1,300–2,000 gigatons per year ( $\text{Gt yr}^{-1}$ )<sup>9</sup>, with giant (longer than 18 km) icebergs representing at least half of the total AIS mass loss<sup>10</sup>. Iceberg trajectory studies for the sea-ice-covered part of the Southern Ocean show that only a small fraction of icebergs calving from the AIS escapes coastal regions directly to the north, owing to topographic steering of ocean currents<sup>9</sup>. Otherwise, a substantial fraction of large icebergs<sup>9</sup> remain entrained in the westward-flowing Antarctic Coastal Current, travelling counter-clockwise around Antarctica before entering the Weddell Sea<sup>11</sup> (Fig. 1). Here, they follow the cyclonic wind-driven Weddell Gyre, where they merge with icebergs calving from the Filchner–Ronne Ice Shelf in the southern Weddell Sea. These icebergs then travel north through the so-called ‘Iceberg Alley’ and exit the Weddell Sea basin to



**Figure 1 | Location map.** Sites MD07-3133 and MD07-3134 are located in the central Scotia Sea. Open arrows refer to Iceberg Alley. The large grey arrow indicates the Southern Hemisphere westerly winds (SHW). The southern boundary of the Antarctic Circumpolar Current (SB of ACC) is indicated by a grey dashed line. White dashed lines describe winter and summer sea-ice extent. Research vessel Polarstern (PS) cores refer to ice-sheet retreat studies conducted in the southeastern Weddell Sea<sup>4</sup>. The inset map (upper right) shows the circum-Antarctic drift of icebergs (turquoise;  $\geq 5$  km in length) calving off the Antarctic ice shelves from 1999 to 2009 (ref. 11). Black arrows indicate the general counterclockwise flow within the Antarctic Coastal Current.

<sup>1</sup>Institute of Geology and Mineralogy, University of Cologne, Zulpicher Strasse 49a, 50935 Cologne, Germany. <sup>2</sup>College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA. <sup>3</sup>Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Am Alten Hafen 26, 27568 Bremerhaven, Germany. <sup>4</sup>International Pacific Research Center, School of Ocean and Earth Sciences and Technology, University of Hawaii at Manoa, 2525 Correa Road, Honolulu, Hawaii 96822, USA. <sup>5</sup>Arctic Centre, University of Lapland, PO Box 122, 96101 Rovaniemi, Finland. <sup>6</sup>Climate Change Research Centre, Level 4, Mathews Building, University of New South Wales, Sydney, New South Wales 2052, Australia. <sup>7</sup>ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, New South Wales 2052, Australia. <sup>8</sup>Hans-Ertel Centre for Weather Research/Climate Monitoring Branch, Meteorological Institute, University of Bonn, Auf dem Hügel 20, 53121 Bonn, Germany.

the northeast into the Scotia Sea (Fig. 1). Around 55° S–60° S, icebergs start to move in a more eastward direction as they are swept into the warmer Antarctic Circumpolar Current. Melt rates remain low until the warmer Antarctic Circumpolar Current is reached, after which the icebergs ablate rapidly in the general area of our core sites.

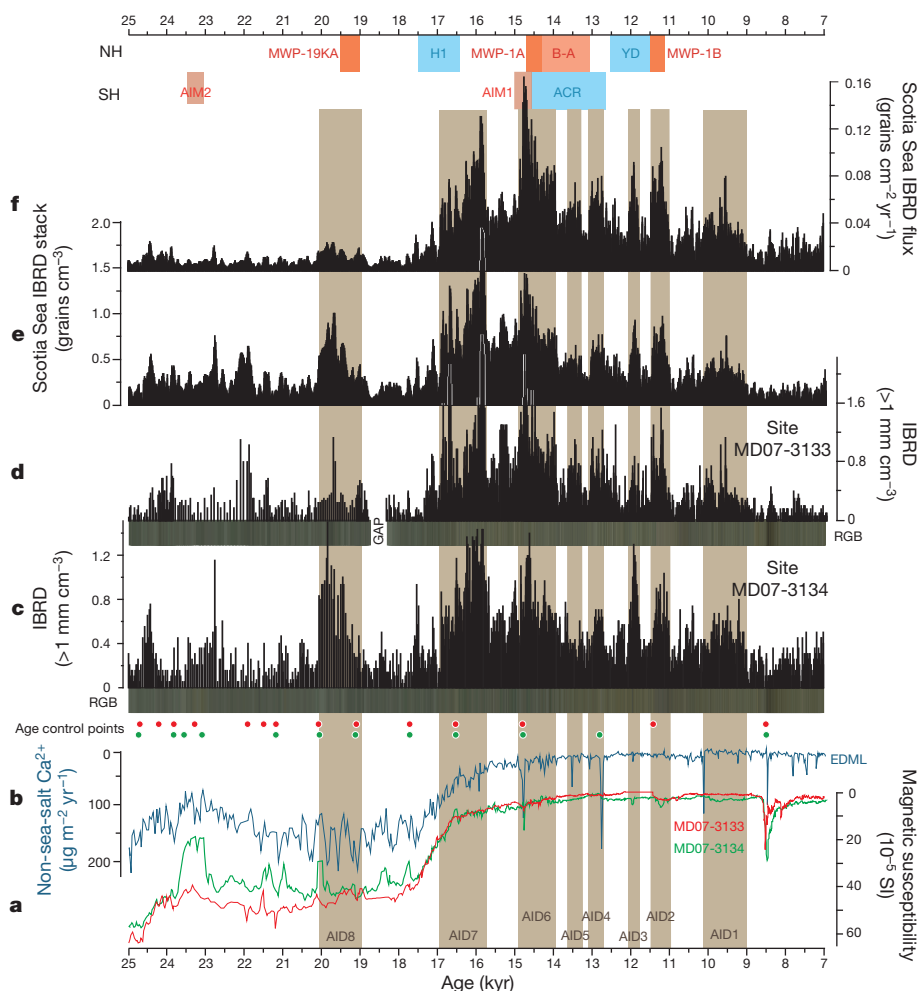
We present two high-resolution records of IBRD from deep-sea core sites MD07-3133 and MD07-3134, collected from the south-central Scotia Sea (Fig. 1) (Methods). Dating relies mainly on the correlation of dust records from our cores to the Antarctic European Project for Ice Coring in Antarctica (EPICA) Dronning Maud Land (EDML) dust record<sup>12</sup> (Methods). Ages reported here refer to the EDML1 or EPICA Dome C (EDC) 3 age scale (Methods). The cores are located in the centre of Iceberg Alley, the primary route taken by large icebergs from the AIS that remained entrained in the Antarctic Coastal Current today as well as during the last deglaciation (Methods). Owing to the loss of some icebergs northwards from the Antarctic Coastal Current, and the gradual ablation of icebergs within the Antarctic Coastal Current, relatively fewer icebergs may reach our sites from more distant sources (Methods). Although the primary IBRD content of icebergs may vary substantially and most of the coarser basal material is released close to the source (that is, the coarse-grained IBRD signal may have a regional bias towards the Weddell Sea and the Antarctic Peninsula), the majority of the fine-grained englacial material determined here should capture a spatially more integrated signal that is representative of variability in the flux of icebergs released by the AIS (except for the Amundsen Sector), particularly in a colder climate than today's.

The IBRD record from site MD07-3134 exhibits a number of peaks starting at about 20,000 years (20 kyr) ago and continuing through the deglaciation until about 9 kyr ago, when modern IBRD levels were reached

(Fig. 2c). Most of the pronounced IBRD events are replicated at site MD07-3133 (Fig. 2d), which is about 300 km to the northwest (Fig. 1), thus providing a robust signal of episodic enhanced iceberg flux through Iceberg Alley. Stacking the IBRD records (Fig. 2e) and their fluxes (Fig. 2f) (Methods) identifies eight primary episodes of enhanced iceberg flux originating from different parts of the AIS that began within one to several decades and lasted from centuries to a millennium. These records thus provide the first evidence for substantial centennial to millennial variability in mass loss from the AIS throughout much of the last deglaciation. We refer to these events as Antarctic Ice Sheet discharges (AID8 to AID1, from oldest to youngest); for error estimates on the timing and duration of AIDs see Methods.

AID8 occurred 20–19 kyr ago (Fig. 2), which corresponds to the start of the deglaciation of the Antarctic Peninsula Ice Sheet<sup>2</sup> and the East AIS in the southeastern Weddell Sea<sup>4</sup>. Although the IBRD event in the stacked record has a concentration of debris similar to that of several of the younger events, its flux (Fig. 2f) is the smallest, which is consistent with an early onset of deglaciation for just the Antarctic Peninsula and Weddell Sea sectors in response to increasing Southern Hemisphere obliquity and sea-level forcing originating from the onset of deglaciation of Northern Hemisphere ice sheets<sup>4</sup>.

Subsequent AID events generally have similar or larger concentration in the IBRD stack than does AID8 (Fig. 2e), whereas the IBRD fluxes of AID7 to AID1 exceed those of AID8 (Fig. 2f). AID7 started 17 kyr ago, with IBRD increasing until around 15.7 kyr ago, when the event abruptly ended. The distinct IBRD peak near the end of AID7 about 16 kyr ago has been documented as a second retreat phase of the Weddell Sea sector of the East AIS<sup>4</sup>. Moreover, most of the marine-based sectors of the Antarctic Peninsula Ice Sheet had retreated by about 16.7 kyr ago<sup>2</sup>, along



**Figure 2 | Climate development from the Last Glacial Maximum to the Holocene (25–7 kyr ago).** Eight phases of enhanced iceberg flux through Iceberg Alley were detected and named AID1 to AID8. **a**, Magnetic susceptibility (in SI units) record of sites MD07-3133 and MD07-3134 (ref. 12). **b**, EDML ice-core non-sea-salt Ca flux on the EDML1/EDC3 timescale<sup>27</sup>. **c**, IBRD with red, green, blue (RGB) colour log for site MD07-3134 and site MD07-3133 (**d**). **e**, Stacked IBRD record. **f**, Stacked IBRD flux record (Methods). AID8–AID1 occurred at approximately 20–19 kyr ago, 17–16 kyr ago, 15–14 kyr ago, 13.5 kyr ago, 13 kyr ago, 12 kyr ago, 11 kyr ago and 10–9 kyr ago. The top panel identifies Northern Hemisphere (NH) climate signals MWP-19KA, Heinrich event 1 (H1), MWP-1A, Bølling–Allerød (B–A), Younger Dryas (YD), and MWP-1B, as well as Southern Hemisphere (SH) Antarctic Isotopic Maxima (AIM) 2 and 1, and the Antarctic Cold Reversal (ACR).

with sectors of the West AIS (for example, the Pine Island glacier of the Amundsen Sea embayment) at 16.4 kyr ago (Methods) that are generally coincident with the major increase in the Scotia Sea IBRD flux rates starting at 16.4 kyr ago (Fig. 2f).

The start of AID7 follows the onset of Antarctic-wide deglacial warming<sup>13</sup> (Fig. 3f) and the retreat of sea ice<sup>13</sup> (Fig. 3e). As indicated by a decrease in the stable carbon isotope ratio of atmospheric carbon  $\delta^{13}\text{C}_{\text{atm}}$  (ref. 14; Fig. 3c), enhanced Southern Ocean opal flux (Fig. 3d), and declining dust/magnetic susceptibility contents (Fig. 2a,b), the onset of AID7 is accompanied by enhanced Southern Ocean ventilation and changes in atmospheric circulation that may have caused an encroachment of warmer water to the AIS marine margins, leading to destabilization and a period of enhanced iceberg release.

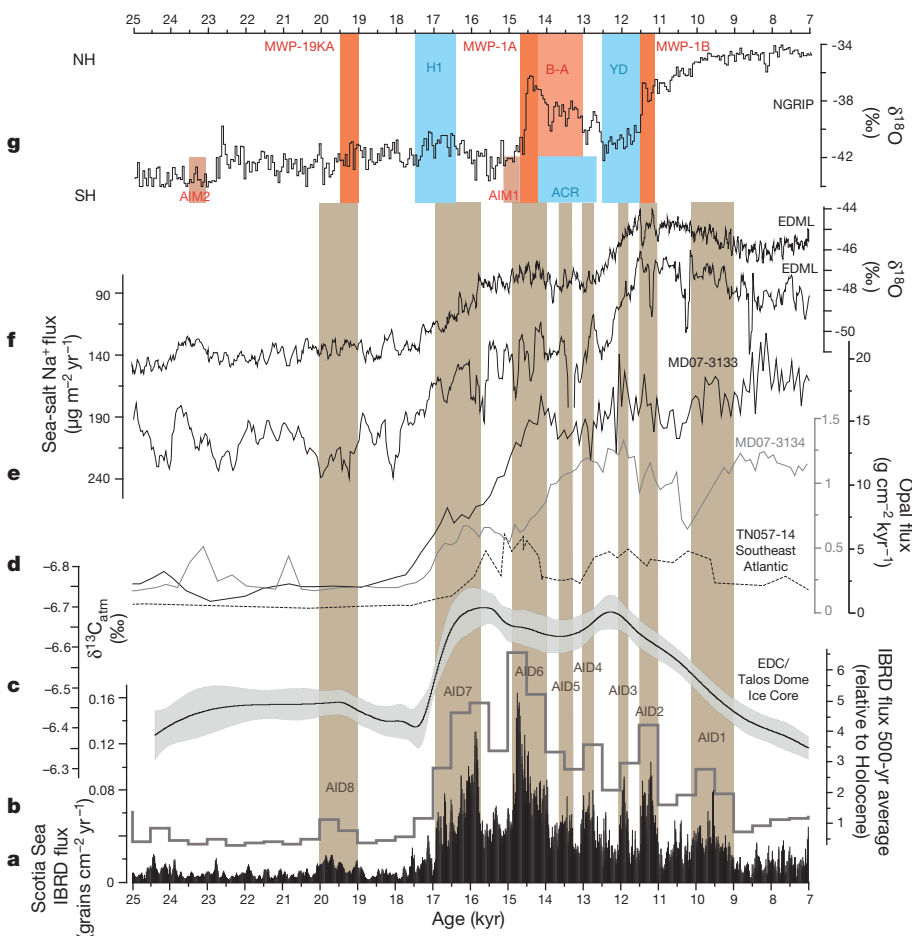
There is a distinct pause in ice rafting between AID7 and AID6 (Figs 2, 3) accompanied by a slowdown in atmospheric warming in the Atlantic sector. The subsequent AID6 exhibits the largest IBRD flux (Fig. 3a and b). AID6 has a relatively abrupt onset at about 15 kyr ago and reaches its peak values approximately 14.8–14.4 kyr ago, followed by a more gradual decline until 13.9 kyr ago, when it abruptly ends. There is general agreement among the few existing well-dated records from Antarctica for widespread retreat of the AIS at this time, including from Mac Robertson Land of the East AIS, the Ross Sea sector of the West AIS, and the Antarctic Peninsula Ice Sheet<sup>2,3</sup> (Methods).

The interval of peak IBRD flux during AID6 (14.8–14.4 kyr ago  $\pm$  0.25 kyr, where the error is  $1\sigma$ ) is synchronous with the onset of the Bølling interstadial in the Northern Hemisphere (14.64  $\pm$  0.09 kyr ago) (Fig. 3g) and with the period of rapid sea-level rise referred to as meltwater pulse (MWP) 1A (occurring 14.65–14.3 kyr ago  $\pm$  0.03 kyr ago)<sup>15</sup>. Fingerprinting studies using the new Tahiti sea-level record and other far-field records<sup>15</sup> support previous such studies<sup>16</sup> in identifying a substantial (at least 50%) Antarctic contribution to the global mean sea-level rise of

about 14 m during MWP-1A. The fact that the peak flux of AID6 in our well-dated IBRD record is the same age and duration (around 400 years) as MWP-1A (as defined from the Tahiti record) provides direct evidence for an Antarctic contribution.

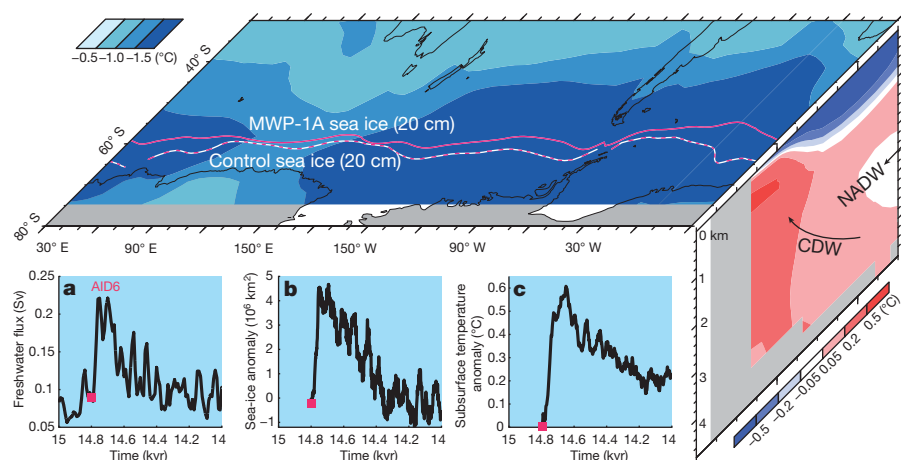
Several factors may have preconditioned or triggered the initial AIS grounding-line (boundary between the floating ice shelf and the grounded ice that rests on bedrock) retreat during AID6, including sea-level rise from melting of Northern Hemisphere ice sheets<sup>17</sup>, subsurface warming associated with changes of the Atlantic Meridional Overturning Circulation<sup>4</sup>, and increases of atmospheric greenhouse gas concentrations. To further explain the climate response to the meltwater discharge associated with AID6, we conducted a series of transient numerical modelling experiments with three different Earth system models (Methods). Assuming that AID6 freshwater forcing corresponded to a maximum sea-level rise of 20 mm yr<sup>-1</sup> (50% of the maximum MWP-1A rate<sup>15</sup>), our model results show considerable Southern Ocean surface cooling, expansion of sea ice, formation of a halocline, reduction of Antarctic Bottom Water formation and subsurface warming at depths of 100–1,500 m through poleward migration of Circumpolar Deep Water (Fig. 4). Consistent with previous modelling experiments<sup>18,19</sup>, we suggest that this ocean thermal forcing accelerated thinning of ice shelves by basal melting, inducing a positive feedback by causing grounding-line retreat, calving and subsequent release of more fresh water.

The next five IBRD events (AID5–AID2), which are of comparable concentration and duration, show a pacing with a repeat time of about 800–900 years. AID2 has the largest IBRD flux of these five events, attaining its peak flux at about 11.3 kyr ago, which corresponds to the age of MWP-1B recorded in the Barbados sea-level record<sup>20</sup>. The absence of an equivalent sea-level event in the Tahiti coral record<sup>21</sup> might reflect different far-field sea-level responses to a deglaciation event from the Pacific sector of the AIS<sup>21</sup>. Although our IBRD record does not distinguish



**Figure 3 | IBRD flux in the Scotia Sea compared to climate changes during the last deglaciation.** a, Stacked IBRD flux record. b, 500-yr averages of stacked IBRD flux relative to Holocene average (Methods). c, Antarctic deglacial  $\delta^{13}\text{C}_{\text{atm}}$  stack with  $1\sigma$  error (shaded)<sup>14</sup>. d, Biogenic opal flux records from the southeast Atlantic site TN057-14 (dashed)<sup>28</sup> and the Scotia Sea sites MD07-3133 (black) and MD07-3134 (grey)<sup>29</sup>. e, EDML ice-core record of sea-salt  $\text{Na}^+$  flux<sup>13</sup>. f,  $\delta^{18}\text{O}$  record from EDML ice core<sup>13</sup>. g,  $\delta^{18}\text{O}$  record from NGRIP ice core<sup>30</sup>. Note that MWP-1A is coeval with AID6, within the dating uncertainties. Note further that major changes in the Southern Hemisphere commenced around 17 kyr ago when the Northern Hemisphere was cold during Heinrich event 1.





**Figure 4 | Three-dimensional pattern of temperature anomalies at 14.8–14 kyr ago (AID6).** The map shows mean surface air temperature anomaly; the vertical profile indicates zonally averaged ocean temperature anomaly in response to an Antarctic meltwater pulse of about 0.22 Sv at around 14.6 kyr ago. The multi-model mean is calculated by averaging the respective anomalies obtained from the Bern3D, LOVECLIM and COSMOS models (Methods). White-and-purple dashed and purple solid lines show the annual mean extent of 20-cm-thick sea ice for 14.8 kyr ago and 14.6 kyr ago, respectively, simulated by the LOVECLIM model. Note that freshwater forcing (a) causes surface cooling, an increase in zonally averaged sea-ice extent (b) and averaged subsurface warming between 800 m and 1,200 m and 63 °S and 70 °S (c). NADW, North Atlantic Deep Water; CDW, Circumpolar Deep Water.

the source of mass loss from the AIS, the presence of AID2 is consistent with an Antarctic contribution to MWP-1B as well<sup>21</sup>.

The youngest IBRD event (AID1) started about 10 kyr ago, with the peak flux occurring around 9.5 kyr ago; it ended about 9 kyr ago. The only other available records that span this time are from the Ross Sea sector of the Western AIS, but they suggest that retreat of the grounding line in the Ross Sea occurred either before or after AID1. The specific sector of the AIS that contributed to this event thus remains unknown (Methods).

Our IBRD record shows clear evidence for episodic mass loss of the AIS during the last deglaciation, but further work is required to identify the sources of mass loss associated with each AID (Methods). Millennial-scale AIDs, similar to those observed in the Northern Hemisphere<sup>5</sup>, are indicative of abrupt ice-sheet instabilities<sup>22</sup> that initiated within one to a few decades, released fresh water into the Southern Ocean, and contributed to the overall deglacial sea-level rise. The peak of AID6 is synchronous, within the dating uncertainties, with MWP-1A. Modelling results indicate that ocean thermal forcing from the transport of warm Circumpolar Deep Water to the base of ice shelves may have been an important positive feedback in AIS deglacial dynamics. Highlighting the role of ocean thermal forcing and rapid ice-sheet responses, our study provides an important palaeoclimate context for contemporaneous<sup>23,24</sup> and future<sup>25,26</sup> interactions between the ice sheet and the ocean around Antarctica.

## METHODS SUMMARY

**Chronology.** Age models for cores MD07-3133 and MD07-3134 (ref. 12) were established by correlating their magnetic susceptibility signal, as well as the Ca and Fe records of MD07-3134, both of which largely reflect dust deposition in the circum-Antarctic region, to the dust (non-sea-salt Ca) flux record of the EDML ice core<sup>27</sup>. We assume that sediment-core magnetic susceptibility, Fe, Ca and ice-core non-sea-salt Ca are in phase because all signals are atmospheric, and therefore essentially reached marine and ice-sheet deposition sites synchronously. Sites MD07-3133 and MD07-3134 have average sedimentation rates of 92 cm kyr<sup>-1</sup> and 63 cm kyr<sup>-1</sup> over the last 36 kyr and 93 kyr, respectively, yielding a deglacial sample resolution for IBRD of 5–15 yr (Methods).

**Sample analysis.** IBRD was counted every centimetre on X-radiographs taken from 1-cm-thick slices that were cut out from the centre of each core segment and exposed to an X-ray system (Methods). Biogenic opal was determined using a combination of a traditional leaching method and newly introduced Fourier transform infrared spectroscopy.

**Earth system modelling experiments.** Coordinated Southern Ocean freshwater perturbation experiments were conducted with the Bern3D and LOVECLIM Earth system models of intermediate complexity and the coupled general circulation model COSMOS (Methods). The IBRD flux stack (Fig. 3a) is used as a Southern Ocean freshwater forcing time series for these models, assuming that 50% of the MWP-1A forcing originated from Antarctica, and attaining peak values of about 0.22 Sv at 14.67 kyr ago, coincident with the onset of the Bølling interstadial (14.64 kyr ago) and MWP-1A (14.65 kyr ago).

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 July 2013; accepted 16 April 2014.

Published online 28 May 2014.

- Clark, P. U. *et al.* The Last Glacial Maximum. *Science* **325**, 710–714 (2009).
- Heroy, D. C. & Anderson, J. B. Radiocarbon constraints on Antarctic Peninsula ice sheet retreat following the Last Glacial Maximum. *Quat. Sci. Rev.* **26**, 3286–3297 (2007).
- Mackintosh, A. *et al.* Retreat of the East Antarctic ice sheet during the last glacial termination. *Nature Geosci.* **4**, 195–202 (2011).
- Weber, M. E. *et al.* Interhemispheric ice-sheet synchronicity during the Last Glacial Maximum. *Science* **334**, 1265–1269 (2011).
- Bond, G. C. & Lotti, R. Iceberg discharges into the North Atlantic on millennial timescales during the last glaciation. *Science* **267**, 1005–1010 (1995).
- Peltier, W. R. Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE. *Annu. Rev. Earth Planet. Sci.* **32**, 111–149 (2004).
- Bentley, M. J. *et al.* Deglacial history of the West Antarctic Ice Sheet in the Weddell Sea embayment: constraints on past ice volume change. *Geology* **38**, 411–414 (2010).
- Conway, H., Hall, B. L., Denton, G. H., Gades, A. M. & Waddington, E. D. Past and future grounding-line retreat of the West Antarctic ice sheet. *Science* **286**, 280–283 (1999).
- Gladstone, R. M., Bigg, G. R. & Nicholls, K. W. Iceberg trajectory modeling and meltwater injection in the Southern Ocean. *J. Geophys. Res.* **106**, 19903–19915 (2001).
- Silva, T. A. M., Bigg, G. R. & Nicholls, K. W. Contribution of giant icebergs to the Southern Ocean freshwater flux. *J. Geophys. Res.* **111**, C03004 (2006).
- Stuart, K. M. & Long, D. G. Tracking large tabular icebergs using the SeaWinds Ku-band microwave scatterometer. *Deep Sea Res. II* **58**, 1285–1300 (2011).
- Weber, M. E. *et al.* Dust transport from Patagonia to Antarctica—a new stratigraphic approach from the Scotia Sea and its implications for the last glacial cycle. *Quat. Sci. Rev.* **36**, 177–188 (2012).
- EPICA Community Members. One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature* **444**, 195–198 (2006).
- Schmitt, J. *et al.* Carbon isotope constraints on the deglacial CO<sub>2</sub> rise from ice cores. *Science* **336**, 711–714 (2012).
- Deschamps, P. *et al.* Ice-sheet collapse and sea-level rise at the Bølling warming 14,600 years ago. *Nature* **483**, 559–564 (2012).
- Clark, P. U., Mitrovica, J. X., Milne, G. A. & Tamisiea, M. E. Sea-level fingerprinting as a direct test for the source of global meltwater pulse 1A. *Science* **295**, 2438–2441 (2002).
- Golledge, N. R., Fogwill, C. J., Mackintosh, A. N. & Buckley, K. M. Dynamics of the Last Glacial Maximum Antarctic ice-sheet and its response to ocean forcing. *Proc. Natl Acad. Sci. USA* **109**, 16052–16056 (2012).
- Menviel, L., Timmermann, A., Timm, O. E. & Mouchet, A. Climate and biogeochemical response to a rapid melting of the West Antarctic ice sheet during interglacials and implications for future climate. *Paleoceanography* **25**, PA4231 (2010).
- Bintanja, R., van Oldenborgh, G. J., Drijfhout, S. S., Wouters, B. & Katsman, C. A. Important role for ocean warming and increased ice-shelf melt in Antarctic sea-ice expansion. *Nature Geosci.* **6**, 376–379 (2013).
- Peltier, W. R. & Fairbanks, R. G. Global glacial ice volume and Last Glacial Maximum duration from an extended Barbados sea level record. *Quat. Sci. Rev.* **25**, 3322–3337 (2006).

21. Bard, E., Hamelin, B. & Delanghe-Sabatier, D. Deglacial meltwater pulse 1B and Younger Dryas sea levels revisited with boreholes at Tahiti. *Science* **327**, 1235–1237 (2010).
22. Marshall, S. J. & Koutnik, M. R. Ice sheet action versus reaction: distinguishing between Heinrich events and Dansgaard-Oeschger cycles in the North Atlantic. *Paleoceanography* **21**, PA2021 (2006).
23. Pritchard, H. D. *et al.* Antarctic ice-sheet loss driven by basal melting of ice shelves. *Nature* **484**, 502–505 (2012).
24. Jenkins, A. *et al.* Observations beneath Pine Island Glacier in West Antarctica and implications for its retreat. *Nature Geosci.* **3**, 468–472 (2010).
25. Hellmer, H. H., Kauker, F., Timmermann, R., Determann, J. & Rae, J. Twenty-first-century warming of a large Antarctic ice-shelf cavity by a redirected coastal current. *Nature* **485**, 225–228 (2012).
26. Gladstone, R. M. *et al.* Calibrated prediction of Pine Island Glacier retreat during the 21st and 22nd centuries with a coupled flowline model. *Earth Planet. Sci. Lett.* **333–334**, 191–199 (2012).
27. Fischer, H. *et al.* Reconstruction of millennial changes in dust emission, transport and regional sea ice coverage using the deep EPICA ice cores from the Atlantic and Indian Ocean sector of Antarctica. *Earth Planet. Sci. Lett.* **260**, 340–354 (2007).
28. Anderson, R. F. *et al.* Wind-driven upwelling in the Southern Ocean and the deglacial rise in atmospheric CO<sub>2</sub>. *Science* **323**, 1443–1448 (2009).
29. Sprenk, D. *et al.* Southern Ocean bioproductivity during the last glacial cycle—new decadal-scale insight from the Scotia Sea. *Geol. Soc. Lond. Spec. Publ.* **381**, 245–261 (2013).
30. NGRIP Members. High-resolution record of Northern Hemisphere climate extending into the last interglacial period. *Nature* **431**, 147–151 (2004).

**Acknowledgements** We acknowledge support from the Deutsche Forschungsgemeinschaft (DFG grant numbers We2039/7-1, Ri525/17-1 and Ku683/

9-1 to M.E.W. and G.K.), the University of Cologne (to M.E.W.), the US NSF Antarctic Glaciology Program (grant numbers ANT-1043517 to P.U.C. and ANT-1341311 to A.T.), the US NSF Paleoclimatology Program and the Japan Agency for Marine-Earth Science and Technology (to A.T.), and Helmholtz funding through the Polar Regions and Coasts in the changing Earth System (PACES) programme (to X.Z., G.L. and G.K.). Our study was also part of the Southern Ocean Initiative of the International Marine Past Global Change Study (IMAGES) program. We thank W. F. Budd for comments on Antarctic ice-sheet dynamics, and M. Winstrop and S. Rasmussen for advice on comparing ice-core chronologies. Experiments with the Bern3D were performed in the Department of Climate and Environmental Physics, University of Bern, and with funding through the Oeschger Center for Climate Change.

**Author Contributions** M.E.W. conceived the idea for the study and, with P.U.C., wrote most of the manuscript. G.K. selected the core sites and provided geochemical data. A.T. oversaw the modelling contributions and helped write the manuscript. R.G. provided insight into iceberg routing and associated ice-sheet modelling. D.S. helped develop the age model and provided biogenic opal data. G.L. and X.Z. contributed results from the COSMOS model. L.M., M.O.C. and T.F. contributed results from Bern3D and LOVECLIM models. C.O. contributed uncertainty estimates on the different age models. All authors commented on the manuscript.

**Author Information** Further data are available at <http://dx.doi.org/10.1594/PANGAEA.819646>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.E.W. ([michael.weber@uni-koeln.de](mailto:michael.weber@uni-koeln.de)).

## METHODS

**Material and sediment composition.** If not tabulated here, data from this study are available at the Pangaea data library (<http://dx.doi.org/10.1594/PANGAEA.819646>). Sediment cores MD07-3133 (57° 26' S, 43° 27' W; 3,101 m water depth; 32.8 m core length) and MD07-3134 (59° 25' S, 41° 28' W; 3,663 m water depth; 58.2 m core length) were obtained from the northern end of Dove Basin and Pirie Bank, respectively, in the southern part of the central Scotia Sea (Fig. 1). Both sites are located well below the carbonate compensation depth, that is, the sediments are largely carbonate-free and consist primarily of biogenic opal and detrital material. The relative proportions of these two components depend on climate conditions (see below): warm periods are associated with homogenous, olive grey to yellowish diatomaceous oozes, whereas cold periods correspond to grey to blue-grey diatom-bearing mud, and climate transitions are associated with olive-grey diatomaceous mud<sup>29</sup>.

**Non-destructive measurements.** We conducted a number of non-destructive measurements<sup>12</sup>. We determined magnetic susceptibility (kappa volume specific) with a Bartington coil sensor for whole-core and a point sensor for split-core measurements at 1-cm increments using a multi-sensor core logger<sup>31</sup>. We also measured the distribution of chemical elements using an AVAATECH X-ray fluorescence core scanner<sup>32</sup>. Results for Ca are given as peak area intensities determined in total counts per second. We performed the analysis using a sample spot of 1 cm × 1 cm in size at 1-cm increments<sup>33</sup>.

We also used a Minolta spectrophotometer CM-2002 to measure  $L^*$ ,  $a^*$  and  $b^*$  as well as RGB colour components<sup>34</sup> at 1-cm increments. RGB colour values were converted to R/G/B<sup>35</sup> and plotted with PanPlot ([www.pangaea.de/software/PanPlot](http://www.pangaea.de/software/PanPlot)) to obtain visual colour logs (Fig. 2).

**Biogenic opal determination.** Biogenic opal was determined by several methods for sites MD07-3133 and MD07-3134 (ref. 29). We first analysed biogenic opal by leaching the sediment in 1 M NaOH solution<sup>36</sup> to obtain ground-truth information at critical transitions (MD07-3133: 1,655–2,785 cm, 112 samples; MD07-3134: 745–1,785 cm, 4,135–4,585 cm; 141 samples). We then implemented a new and fast technique that relies on this ground-truthing, called Fourier transform infrared spectroscopy<sup>37</sup>, to obtain information on biogenic opal at 10-cm increments over the entire core lengths of sites MD07-3133 and MD07-3134 (316 and 575 samples, respectively). Note that biogenic opal flux rates in Fig. 3d are normalized for <sup>230</sup>Th to account for sediment focusing with the exception of MD07-3133, where no <sup>230</sup>Th data have been obtained so far<sup>29</sup>.

**Chronology.** The low-resolution chronology of sites MD07-3133 and MD07-3134, which extend back to 36 kyr ago and 93 kyr ago, respectively, is described in ref. 12. It is based on several ground-truth data points associated with Marine Isotopic Stage boundaries, the correlation to rather massive layers rich in coarser-grained detritus<sup>38</sup>, and the identification of the Laschamp event. The high-resolution chronology relies on dust–climate couplings between Southern Ocean sediment and the Antarctic EDML ice core<sup>12</sup>. The use of marine Fe records as a proxy for dust is a well-established approach reflecting coherent and synchronous changes in dust deposition across much of the Southern Ocean and the AIS, and has recently also been implemented to construct age models for records from the Atlantic<sup>39</sup> and the Pacific<sup>40</sup> sectors of the Southern Ocean to reconstruct Southern Hemisphere dust climate couplings during the Late Quaternary using Fe counts. In addition, a number of previous workers have identified a remarkable similarity in the structure of variability of magnetic susceptibility in cores from the Scotia Sea and the variability in dust records from Antarctic ice cores<sup>41–43</sup>. Pugh *et al.*<sup>44</sup> noted this similarity in magnetic susceptibility records from three cores in the Scotia Sea (that is, the same region as our records cover), and explicitly tested whether the magnetic susceptibility signal was synchronous with the EPICA ice-core dust record. They first established two independent age models for their cores from (1) biostratigraphy (variations in *Cycladophora davisiana* abundance tied to the LR04 time scale  $\delta^{18}\text{O}$  global stack<sup>45</sup>) and (2) radiocarbon dating, and then compared those age models to one constructed by synchronizing their magnetic susceptibility records to the ice-core dust record. They concluded that these three independently derived age models are “mutually consistent over their common ranges” and that millennial-scale fluctuations in magnetic susceptibility and in Antarctic dust concentration were synchronous, supporting our strategy of tuning Scotia Sea magnetic susceptibility records to Antarctic ice-core dust records to develop an age model for our marine records<sup>12</sup>.

Weber *et al.*<sup>12</sup> also noted a remarkable similarity in the structure of variability of Ca in the cores from the Scotia Sea and the variability in the non-sea-salt Ca records from Antarctic ice cores. Weber *et al.*<sup>12</sup> measured Ca by X-ray fluorescence which, because the Scotia Sea sites do not contain biogenic Ca, is the same proxy for dust as used in the ice cores (non-sea-salt Ca), and found that when Ca is tuned to non-sea-salt Ca (which is completely independent of the magnetic susceptibility tuning), the resulting correlation between the two is as strong as the correlation of magnetic susceptibility to non-sea-salt Ca, providing additional support for synchronicity,

because both marine and ice-core Ca records are proxies for dust, with the most likely physical process being short-lived atmospheric transport.

Given these independent lines of evidence supporting synchronous changes, we tuned magnetic susceptibility and Ca in our cores to the non-sea-salt Ca flux record from the EDML ice core<sup>27</sup>. We used three statistical approaches to produce a tuned, high-resolution age model, evaluate its quality and provide error estimates<sup>12</sup>. Implementing dynamic time wrapping algorithms, for instance, helped to detect circularities by computing a time (depth) axis stretch and mapping two time series optimally by minimizing the cumulative distance without any pre-defined tie points. The resulting age–depth structure obtained from dynamic time wrapping was virtually identical to the one obtained by the tuning procedure, which served as independent verification of our final tuning with tie points.

Our deglacial age model is based primarily on five stratigraphic tie points (TP1 to TP5) that are shared between the records and reveal a detailed record for the period about 20–8 kyr ago over 10 m and 7 m core lengths for sites MD07-3133 and MD07-3134, respectively (Extended Data Fig. 1). The five tie points correspond to consistent breaks in slopes and characteristic and reproducible peaks in the records of magnetic susceptibility and Ca relative to non-sea-salt Ca. Our deglacial chronology also includes Fe, which shows the same characteristic trends and features as magnetic susceptibility and Ca, and which has been used to synchronize Southern Ocean marine sediment and Antarctic ice-core records<sup>39,40</sup>. TP1 (about 19 kyr ago) is one of many correlation points for the Last Glacial Maximum used in ref. 12. It marks the end of distinctly elevated values in magnetic susceptibility, Ca, Fe and non-sea-salt Ca towards the end of this phase. TP2 (about 17.6 kyr ago) is defined on the basis of a steep increase in slope in each of the records. A subsequent decrease in slope in each of the records is the basis for TP3 around 16.5 kyr ago. TP4 occurs another 1,800 years later with another distinct decrease in slope in each record around 14.7 kyr ago and a corresponding distinct peak in the non-sea-salt Ca and magnetic susceptibility records. TP5 is recorded in all archives as a distinct peak that occurred around 8.4 kyr ago.

In addition to these common tie points, site MD07-3134 and EDML share a dust peak at 12.85 kyr ago that is not preserved at site MD07-3133. Sites MD07-3133 and MD07-3134 share a weak dust peak at 11.42 kyr ago, which was taken as an additional age control for MD07-3133. This correlation is supported by the fact that colour component  $a^*$  (red–green variability) shows a distinct peak at 8.5–8.3 kyr ago at sites MD07-3133 and MD07-3134, which also serves as control point at both sites (Fig. 2c, d). This peak is replicated at 11.42 kyr ago (that is, at 19.00 m at site MD07-3133, and at 11.00 m at site MD07-3134).

Based on our age model, sedimentation rates in our cores are high, 0.2–1.2 m kyr<sup>−1</sup> for site MD07-3134 (63 cm kyr<sup>−1</sup> on average over the last 93 kyr) and 0.3–2.1 m kyr<sup>−1</sup> for site MD07-3133 (92 cm kyr<sup>−1</sup> on average over the last 36 kyr), translating into mass accumulation rates of 10–120 g cm<sup>2</sup> kyr<sup>−1</sup> and 15–170 g cm<sup>2</sup> kyr<sup>−1</sup>, and into sample resolutions of 8–50 yr and 5–33 yr (given a sample increment of 1 cm) for IBRD counting, respectively.

Considering the chronological limitations and problems that are associated with most Southern Ocean sediment archives, the age model of our two Scotia Sea sites has a high resolution and should be very accurate because the uncertainties associated with <sup>14</sup>C dating, such as reservoir age, are circumvented. Therefore, our expanded deglacial record of about 20–8 kyr ago (documented over 7 m and 10 m core length in sites MD07-3134 and MD07-3133, respectively) is chronologically better constrained than the deglacial signals that have been inferred from shallow-water archives around the Antarctic continent, such as at Mac Robertson Land<sup>3</sup>, where inferred changes usually occur over a few centimetres in condensed, lithologically complex successions with partially reversed <sup>14</sup>C ages that have uncertain reservoir ages.

**Age scale and AID uncertainty estimates.** We use the EDML1/EDC3 age model that was developed in conjunction with the EDC age model<sup>27</sup> by incorporating ground-truth stratigraphic evidence, most importantly the correlation of regional volcanic ash layers. Based on this age model, three AIDs in the Scotia Sea correlate well with large-scale events: the peak of AID7 with the major retreat of the Weddell Sea part of the East AIS at 16.1 kyr ago<sup>4</sup>, the peak of AID6 with MWP-1A around 14.7 kyr ago, and AID2 with MWP-1B at 11.3 kyr ago.

The original age scale EDML1/EDC3 relied on ash correlation to EDC<sup>46,47</sup>. The errors of this correlation are a couple of decades. For ages younger than 41 kyr ago, the EDML1/EDC3 age model is based on correlation of methane and <sup>10</sup>Be to the Greenland Ice Core Chronology 2005 (GICC05), which largely relies on layer counting<sup>48</sup>. The errors of GICC05 were also a couple of decades for the Holocene, and up to two centuries for the time around MWP-1A. Subsequent revision to the EDML age model<sup>49</sup> and the latest Antarctic-wide ice-core chronology<sup>50</sup>, called AICC2012, tried to account for regional stratigraphic constraints and the need to adjust ages to achieve interhemispheric methane and <sup>10</sup>Be synchronization during deglaciation.

AICC2012 shows little to no difference to EDML1/EDC3 for most of the Holocene: that is, the timing of AID1 and AID2 are not affected. However, from about 8–18 kyr ago, AICC2012 becomes progressively older, with an age difference peaking



at about 550 yr at 18 kyr ago, and declining again to about 200 yr from the Last Glacial Maximum. For the time of MWP-1A, for instance, the age difference is almost 400 yr. This would increase the age of the tie point, which occurs right at the start of AIE6, from 14.8 kyr ago to almost 15.2 kyr ago, implying that the peak of AIE6 dates 15.1–14.7 kyr ago instead of 14.8–14.4 kyr ago. The entire AIE6 would hence be 15.2–14.2 kyr ago.

However, this age shift is only achieved at the expense of a significantly larger dating uncertainty than the original EDML1/EDC3 chronology. For the Holocene the uncertainty is around 200 yr (ref. 51). At about 12 kyr ago, it increases to almost 500 yr, and it reaches 1,000 yr by about 19 kyr ago. The original EDML1/EDC3 age scale, however, has much lower uncertainties. From approximately 10 kyr ago to 14 kyr ago, uncertainties increase from only a few decades to around 200 yr, respectively (Extended Data Fig. 2d), and then further increase to about 700 yr at around 18 kyr ago. One should note that, although AICC2012 is applied to all Antarctic ice cores, the only ground-truth correlation of EDML is established via volcanics to EDC.

Synchronization of the new West AIS ice core to EDML via layer counting between volcanic acidic horizons, which has not been included in AICC2012, indicates that AICC2012 is significantly too old for the pre-Holocene time (Winstrup, M., Vinther, B. M., Sigl, M., McConnell, J., Svensson, A. M. and Wegner, A., manuscript in preparation). For the depth around MWP-1A, for instance (about 800 m at EDML), this new chronology will again be closer to the timing of the original EDML1/EDC3 timescale. Also, the uncertainty will again be much lower, given the low counting error of annual bands between volcanic ashes.

Since the ageing of AICC2012 is only achieved at the expense of much higher uncertainties relative to the EDML1/EDC3 age scale, and all shifts are within the uncertainties of the different ice-core age models, we use the original EDML1/EDC3 age model. We note also that its dating uncertainties are lower and that it has more ground-truth chronological constraints.

The original uncertainty estimates of the Scotia Sea sites<sup>12</sup> built upon a bootstrap algorithm that simulates possibly wrong tie points and therefore provides conservative error estimates. However, these uncertainty estimates do not include ice-core age uncertainties. We used the resulting dating errors ( $2\sigma$ ) of the tie point correlation to calculate uncertainties for the start, end and duration of AIDs (Extended Data Fig. 2a). In addition, we considered the uncertainties in the EDML1 chronology<sup>47</sup> as well as the EDC3 chronology<sup>46</sup> when calculating the full uncertainties for the AIDs (Extended Data Fig. 2b). All uncertainties are incorporated into the absolute dating uncertainty ( $2\sigma$  error) of the AIDs (Extended Data Table 1). The bootstrap algorithm allows for the derivation of relative dating uncertainties for the AIDs. Since interpolation errors of nearby layers are highly correlated, these uncertainties are naturally smaller than the absolute dating uncertainties (Extended Data Fig. 2c). Hence, significantly more precise estimates can be made on the duration of and the distance between AIDs.

**IBRD counting and conversion.** We cut out 1-cm-thick, 10-cm-wide and 25-cm-long slices from the centre of each core section, placed them on plastic plates and exposed the slices to film-sensitive negatives for 3 min to 5 min using a HP 43855 X-ray system. The negatives were then scanned at 300 dpi (dots per inch) and analysed for the amount of IBRD using a 1-mm by 1-mm grid.

We counted all visible grains  $> 1$  mm in diameter per centimetre core depth<sup>52</sup> (that is, per  $10\text{ cm}^3$  volume) as an indicator for the content of IBRD (Extended Data Fig. 3). We distinguished between grains 1–2 mm and  $> 2$  mm in diameter. Also, to account for operator subjectivity in IBRD counts, three different individuals conducted independent counts, which were then averaged for Fig. 2c and d. For Fig. 2e, we stacked the results by re-sampling each data series at 10-yr increments, adding the individual counts from the two cores, and then dividing them by 2. For Fig. 2f, we converted the IBRD time series from sites MD07-3133 and MD07-1334 into flux rates, using the detailed age models<sup>12</sup>, and, again, stacked the results. For Fig. 3b, we averaged the stacked IBRD flux rates into 500-yr slices.

The stacked IBRD flux-rate record indicates, on average, rather stable values with little multi-centennial variability of  $0.01\text{ grains cm}^2\text{ yr}^{-1}$  for 25–21 kyr ago, and  $0.015\text{ grains cm}^2\text{ yr}^{-1}$  for 8–0 kyr ago. We infer that these low values represent steady-state conditions, that is, the amount of iceberg calving off the Antarctic continent is balanced by accumulation on the ice sheet. During deglaciation, however, flux rates were distinctly above these low values, and 10–15 times higher during meltwater pulses. Even when averaged into 500-yr slices (grey histogram line, Fig. 3e), meltwater pulses show IBRD flux rates 4–7 times higher.

We note that icebergs usually lose the majority of their coarse-grained IBRD (pebbles to boulders) immediately after calving proximal to the AIS grounding lines. The largest IBRD reaching our distal core positions in Iceberg Alley is usually 1–3 mm in size and dispersed in a diatom-rich or clay-rich matrix (Extended Data Fig. 3). These particles embedded within icebergs can travel over long distances and argue against a strong regional bias in our IBRD data set.

**On the relation of AIS mass loss, icebergs and IBRD.** The majority of IBRD originates from bedrock, either directly (through scouring) or indirectly (through

refreezing of sediment-laden water at the bed), and is entrained in the deeper part of a glacier. Where glaciers merge, medial moraines can cause IBRD to occur over a much greater thickness, sometimes even throughout the ice column (W. F. Budd, personal communication, 2013). Other processes internal to ice flow (for example, compressive flow) can also transport debris from the bed well up into the ice column. Only a small fraction will be delivered onto the ice surface, for example, by rock fall, or by wind transport in the case of finer grains.

In the floating ice shelves most IBRD near the base will melt out relatively close to the source<sup>53,54</sup>, immediately in front of the grounding line. In ice shelves with high melt rates, the whole sediment-rich basal layer might ablate. However, in some ice shelves tributary glaciers can carry IBRD straight into a basal freeze-on zone, where they are protected from melting out by a layer of meteoric ice (for example, the west side of the Amery; see figures 6a and 9a of ref. 55). This process and the potential for localized IBRD throughout the ice column allow some IBRD to reach the front of large ice shelves.

The fact that we capture mostly small, englacial IBRD in Iceberg Alley indicates that the majority of the coarse-grained IBRD has been deposited before the icebergs reached the Scotia Sea sites. The fact that englacial material can travel long distances and will probably survive the counterclockwise journey around Antarctica (see below) within the Antarctic Coastal Current could therefore indicate that regional bias is not strong and no large nonlinearities exist between the amount of entrained IBRD detected in the Scotia Sea and the number of icebergs calving off the AIS.

**Iceberg Alley today and during deglaciation.** The term ‘Iceberg Alley’<sup>56</sup> refers to the confluence zone where Antarctic icebergs merge<sup>57</sup> and exit the Weddell Sea to the north into the Scotia Sea<sup>58</sup>. Icebergs take several months to two years to complete their journey counterclockwise around Antarctica<sup>9</sup>. Although icebergs lose some mass in transit (through wave erosion, friction or collisions with sea ice), it is a rapid journey in cold waters. Therefore, melt rates remain low until the warmer Antarctic Circumpolar Current is reached, after which the icebergs ablate rapidly<sup>59</sup>. This is one of the reasons why the sites from Iceberg Alley should contain a circum-Antarctic-wide IBRD signal, although it is also reasonable to assume that the Weddell Sea part of the East AIS and the Antarctica Peninsula Ice Sheet provide more IBRD than more distal sources. This regional bias has been detected before in IBRD provenance studies<sup>53,54</sup>, although deglaciations may involve different iceberg fluxes from various sources at different times as well as longer transport distances due to colder climates<sup>60</sup>.

A substantial percentage of large icebergs reach Iceberg Alley through the Antarctic Coastal Current<sup>11</sup>. Some of the smaller icebergs (for example, in the vicinity of the Kerguelen plateau) may stray directly north due to topographic steering of ocean currents, but even the smaller icebergs mostly stay entrained in the Antarctic Coastal Current. Most icebergs calving from the Western AIS are not likely to reach Iceberg Alley through the Antarctic Coastal Current because there is a small northwards and clockwise escape route from the coastal drift (see arrow in Fig. 1; refs 9, 11). However, it is conceivable that these icebergs could also come through the Scotia Sea and affect the IBRD record of our sites.

Several factors indicate that the present-day iceberg routing through Iceberg Alley would have remained largely the same during the last deglaciation. The Antarctic Divergence is the major driver of the ocean–atmospheric circulation in the Southern Ocean. It forces winds north of it to blow east and to feed clockwise into the Antarctic Circumpolar Current. South of the Antarctic Divergence, winds blow west and feed counterclockwise into the Antarctic Coastal Current around Antarctica. The Antarctic Circumpolar Current is constrained to flow south of the tip of South America at  $56^\circ\text{ S}$  as it passes from the Pacific Ocean to the Atlantic Ocean, and its mean position therefore lies between  $50^\circ\text{ S}$  and  $55^\circ\text{ S}$  (ref. 61). The Antarctic Circumpolar Current and Antarctic Coastal Current will thus exist under different climates of the past<sup>28,62</sup>, although specific fronts associated with them probably shifted north or south, depending on the state of the climate and its associated sea-ice coverage.

An additional factor that is important in routing of icebergs discharged from the AIS is the Coriolis force, which depends on the rotation of the Earth and is therefore independent of the state of the climate. It causes all moving objects to be deflected to the left in the Southern Hemisphere. Thus, all icebergs calving from the AIS are forced to move counterclockwise and maintain a course parallel to the edge of the Antarctic continent. Moreover, the Coriolis force exerts a greater force on the larger icebergs<sup>57</sup>, which contribute most of the AIS mass loss, and they thus largely remain entrained in the Antarctic Coastal Current (that is, sea ice does not have a large effect on altering the course of big icebergs); preferably smaller icebergs and crawlers would have escaped the Antarctic Coastal Current during times of thicker sea-ice coverage further away from the continent.

Finally, several studies<sup>4,63</sup> have reconstructed extensive coastal polynyas (areas of open water surrounded by sea ice) immediately in front of the AIS during colder periods when sea-ice coverage was dense further away from the continent. This further implies a vigorous Antarctic Coastal Current also during glacial and deglacial times. In summary, there are several lines of evidence that iceberg routing through

Iceberg Alley operated in much the same way during deglaciation as is observed and modelled today.

We note that Kanfoush *et al.*<sup>38</sup> detected layers rich in coarser-grained detritus across the southern Atlantic, and named them SA0 (about 14–15 kyr ago) to SA6 (about 55 kyr ago). Whereas they interpreted these layers as IBRD and also inferred episodes of Antarctic ice-sheet instability from them, subsequent studies<sup>64</sup> demonstrated that the material was volcanic ash, most probably from the South Sandwich Islands, and concluded that the ash was deposited on sea ice, which then transported the material. In any event, the majority of Antarctic icebergs are entrained into the Antarctic Circumpolar Current in Iceberg Alley and melt long before these distal South Atlantic sites can be reached<sup>57</sup>. Moreover, the layers described by Kanfoush *et al.*<sup>38</sup> are distinct and massive, whereas our IBRD counts usually indicate small (1–3 mm in diameter) and dispersed grains embedded in a diatom- or clay-rich matrix. In addition, our IBRD records show no long-term correlation to the sea-salt Na record (Fig. 3f) of the EDM1 ice core, a proxy for sea ice. Accordingly, we are confident that the Scotia Sea IBRD counts indicate a source from icebergs with varying fluxes rather than a fallout from volcanic eruptions and transport by sea ice.

**Deglacial meltwater sources.** During deglaciation, specific parts of the AIS contributed to specific retreat events, that is, ice-sheet retreat probably did not occur contemporaneously across the entire AIS. The Weddell Sea sector of the East AIS probably contributed to AID8 (MWP-19KA; ref. 1) and AID7, with evidence that ice-sheet retreat occurred 19 kyr ago and 16 kyr ago, respectively<sup>4</sup>. Early deglaciation of the Antarctic Peninsula<sup>65</sup> and the Lambert glacier–Amery ice shelf system<sup>66</sup> may also have contributed icebergs to these events. In addition, the Pine Island glacier of the Amundsen Sea embayment retreated at 16.4 kyr ago<sup>67</sup>, and may have contributed to AID7.

Evidence for widespread retreat 14–15 kyr ago comes from additional sectors of the East AIS, that is, Mac Robertson Land<sup>3</sup>, as well as the Ross Sea sector of the West AIS<sup>68,69</sup> and the Antarctic Peninsula Ice Sheet<sup>2,70</sup>, suggesting multiple sources for AID6 (MWP-1A). Such a change in the distribution of meltwater sources is supported by our freshwater experiments, which simulate increased subsurface temperature responses when sources are shifted from west to east during deglaciation. Subsequent retreat during AID2 (MWP-1B) is recorded at Wilkes Land, part of the East AIS<sup>71</sup>. Available terrestrial and shallow-marine records from the Ross Sea sector of the West AIS suggest retreat of the grounding line in the Ross Sea as having occurred either before<sup>72</sup> or after<sup>8</sup> AID1. Therefore, the specific sector of the AIS that contributed to this event remains unknown.

**Ice-sheet/ocean feedbacks during AID6.** To study the response of the Southern Hemisphere climate system to AID6-related iceberg surges, a series of coordinated transient climate modelling experiments is conducted with a hierarchy of climate models using time-varying Southern Ocean freshwater forcing (Extended Data Table 1). The freshwater forcing, which is evenly applied to the climate models in the Weddell and Ross seas, is a linearly scaled version of the IBRD flux stack of Fig. 3a. The scaling factor is determined by assuming that 50% of the maximum sea level rates during MWP-1A ( $40 \text{ mm yr}^{-1}$ )<sup>15</sup> originated from Antarctica. The corresponding peak freshwater forcing around Antarctica during AID6 reaches values of 0.22 Sv (Fig. 4). It should be noted here that the anomalous fresh water released from the Ross and Weddell seas is picked up by the Antarctic Circumpolar Current. This results in an efficient mixing of the freshwater around Antarctica, mimicking to some degree the effect of drifting icebergs<sup>10</sup> and their corresponding freshwater fluxes. The multi-model approach allows us to determine the robust features of the climate response to an Antarctic iceberg discharge 14.8–14.0 kyr ago (Fig. 3).

The following three climate models have been used: (1) The Bern3D model<sup>73,74</sup> is an intermediate-complexity model based on an energy-balance model for the atmosphere and a frictional geostrophic model for the ocean. Other components included in the Bern3D model, such as a carbon cycle and sediment component, are not considered here. The horizontal resolution adopted here is  $5^\circ$  by  $10^\circ$ . Atmospheric winds are prescribed at present-day values, thus inhibiting large-scale atmospheric circulation changes in response to the climate change. The freshwater forcing is applied to the Bern3D model under climate boundary conditions that represent the Last Glacial Maximum. (2) The global climate model LOVECLIM uses a simplified dynamical atmosphere coupled to an ocean general circulation model, and dynamic/thermodynamic sea-ice and diagnostic vegetation models. It has been used in a number of palaeoclimate studies focusing on the Southern Ocean<sup>18,75,76</sup>. Horizontal resolutions of the atmosphere and the ocean are  $5.6^\circ$  by  $5.6^\circ$  and  $3^\circ$  by  $3^\circ$ , respectively. Here the Southern Ocean freshwater forcing is applied to constant climate boundary conditions that represent the climate around 15 kyr ago<sup>77</sup>. These initial conditions were obtained from 5,000-yr-long spin-up experiments that use 15-kyr-ago greenhouse gas concentrations, orbital forcing and ice-sheet forcing. (3) COSMOS is a comprehensive coupled general circulation model including a land surface scheme and a dynamic sea-ice model. It has been extensively used to study various key palaeoclimate periods<sup>78–80</sup>. Horizontal resolutions of the atmosphere

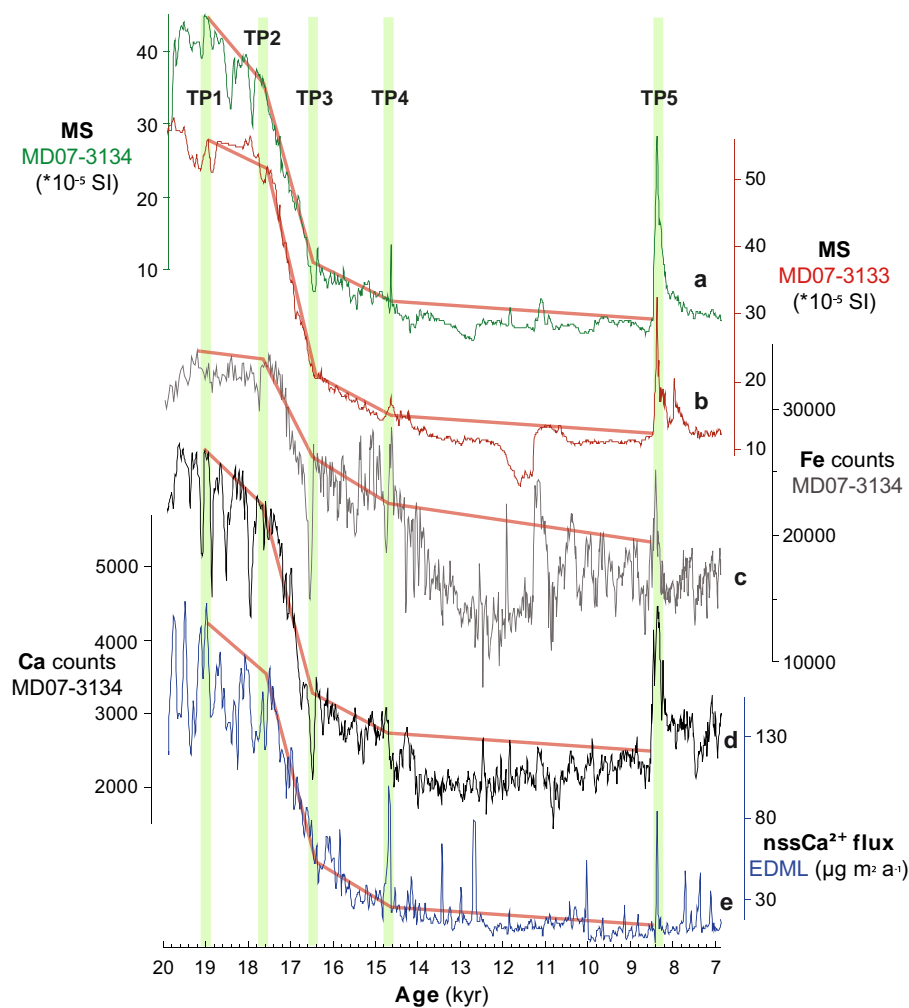
and the ocean are  $3.75^\circ$  and  $3^\circ$ , respectively. The freshwater forcing is applied under 15-kyr-ago boundary conditions. The initial state for the freshwater forcing experiment was obtained from a transient deglacial simulation varying  $\text{CO}_2$  concentrations derived from EDC ice core<sup>81</sup> and orbital forcing.

Mimicking the effects of AID6, the applied Southern Hemisphere freshwater forcing causes widespread surface cooling in the Southern Hemisphere and an increase of Southern Ocean sea-ice area (Fig. 4), in accordance with the sea-salt  $\text{Na}^+$  flux data from EPICA Dronning Maud Land (Fig. 3e). This cooling would contribute to establishing the Antarctic Cold Reversal and the bipolar seesaw response during the Bølling–Allerød. Moreover, the freshwater forcing generates a strong halocline, which inhibits formation of very cold Antarctic Bottom Water, which is then replaced by considerably warmer Circumpolar Deep Water (Fig. 4). Following previous studies<sup>18,19</sup>, we hypothesize that the associated sub-surface warming at depths of 100–1,500 m can cause ice-shelf melting, grounding-line retreat, calving and subsequent release of more fresh water, all of which act as positive feedbacks.

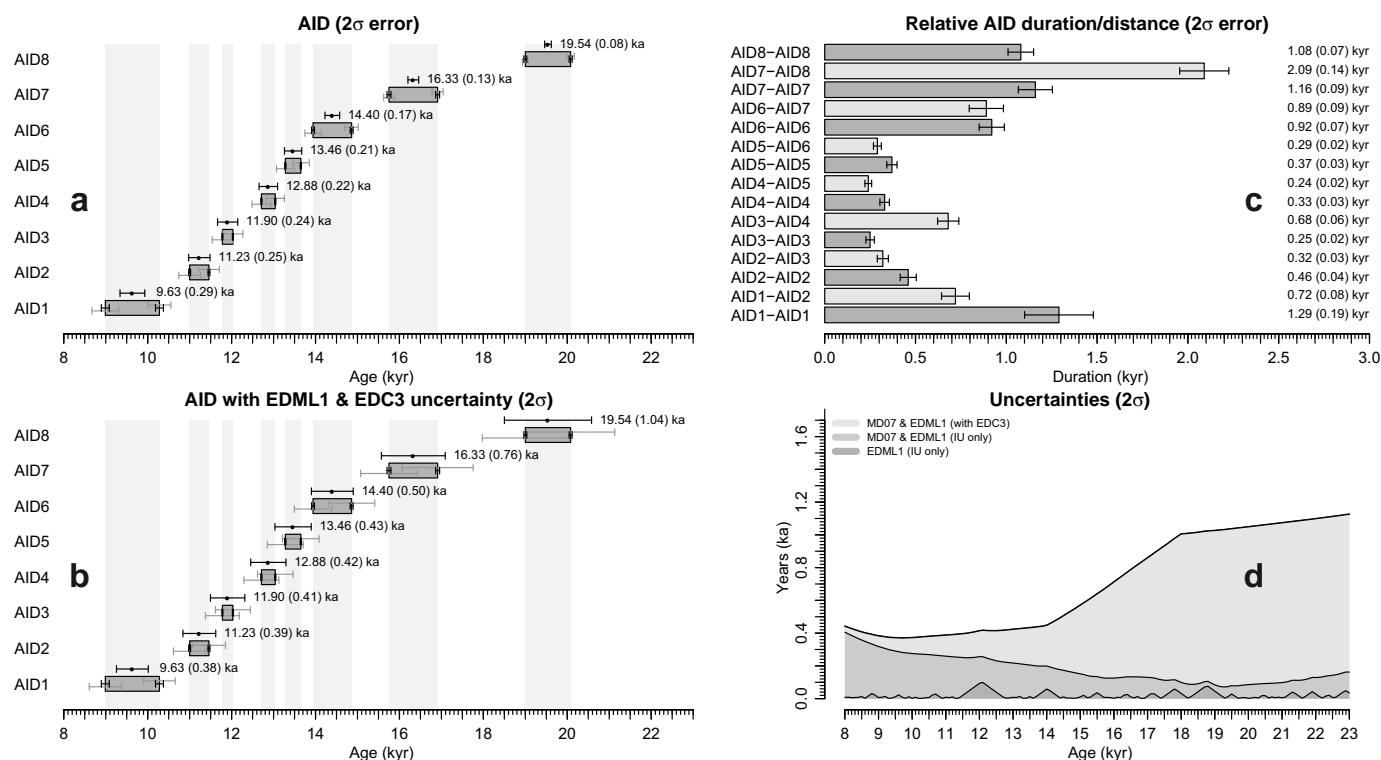
- Weber, M. E., Niessen, F., Kuhn, G. & Wiedicke, M. Calibration and application of marine sedimentary physical properties using a multi-sensor core logger. *Mar. Geol.* **136**, 151–172 (1997).
- Jansen, J. H. F., Van der Gaast, S. J., Koster, B. & Vaars, A. J. CORTEX, a shipboard XRF-scanner for element analyses in split sediment cores. *Mar. Geol.* **151**, 143–153 (1998).
- Richter, T. O. *et al.* The Avaatech XRF Core Scanner: technical description and applications to NE Atlantic sediments. *Geol. Soc. Lond. Spec. Publ.* **267**, 39–50 (2006).
- Weber, M. E. Estimation of biogenic carbonate and opal by continuous non-destructive measurements in deep-sea sediments: application to the eastern Equatorial Pacific. *Deep Sea Res.* **45**, 1955–1975 (1998).
- Weber, M. E. *et al.* BMPix and PEAK tools: new methods for automated laminae recognition and counting—application to glacial varves from Antarctic marine sediment. *Geochim. Geophys. Geosyst.* **11**, 1–18 (2010).
- Müller, P. J. & Schneider, R. An automated leaching method for the determination of opal in sediments and particulate matter. *Deep Sea Res.* **40**, 425–444 (1993).
- Rosen, P. *et al.* Fourier transform infrared spectroscopy, a new method for rapid determination of total organic and inorganic carbon and biogenic silica concentration in lake sediments. *J. Paleolimnol.* **43**, 247–259 (2010).
- Kanfoush, S. I. *et al.* Millennial-scale instability of the Antarctic Ice Sheet during the last glaciation. *Science* **288**, 1815–1819 (2000).
- Martínez-García, A. *et al.* Southern Ocean dust-climate coupling over the past four million years. *Nature* **476**, 312–315 (2011).
- Lamy, F. *et al.* Increased dust deposition in the Pacific Southern Ocean during glacial periods. *Science* **343**, 403–407 (2014).
- Hofmann, A. *Kurzfristige Klimaschwankungen im Scotiameer und Ergebnisse zur Kalbungsgeschichte der Antarktis während der letzten 200 000 Jahre (Rapid climate oscillations in the Scotia Sea and results of the calving history of Antarctica during the last 200 000 years)* PhD thesis (Univ. Bremen, 1999).
- Diekmann, B. *et al.* Terrigenous sediment supply in the Scotia Sea (Southern Ocean): response to Late Quaternary ice dynamics in Patagonia and on the Antarctic Peninsula. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **162**, 357–387 (2000).
- Yoon, H. I., Khim, B. K., Yoo, K. C., Bak, Y. S. & Lee, J. I. Late glacial to Holocene climatic and oceanographic record of sediment facies from the South Scotia Sea off the northern Antarctic Peninsula. *Deep Sea Res.* **54**, 2367–2387 (2007).
- Pugh, R. S., McCave, I. N., Hillenbrand, C. D. & Kuhn, G. Circum-Antarctic age modelling of Quaternary marine cores under the Antarctic Circumpolar Current: ice-core dust–magnetic correlation. *Earth Planet. Sci. Lett.* **284**, 113–123 (2009).
- Lisiecki, L. E. & Raymo, M. E. A Pliocene–Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* **20**, PA1003 (2005).
- Parrenin, F. *et al.* The EDC3 chronology for the EPICA Dome C ice core. *Clim. Past* **3**, 485–497 (2007).
- Ruth, U. *et al.* “EDML1”: a chronology for the EPICA deep ice core from Dronning Maud Land, Antarctica, over the last 150 000 years. *Clim. Past* **3**, 475–484 (2007).
- Rasmussen, S. O. *et al.* A new Greenland ice core chronology for the last glacial termination. *J. Geophys. Res.* **111**, D06102 (2006).
- Lemieux-Dudon, B. *et al.* Consistent dating for Antarctic and Greenland ice cores. *Quat. Sci. Rev.* **29**, 8–20 (2010).
- Veres, D. *et al.* The Antarctic ice core chronology (AICC2012): an optimized multi-parameter and multi-site dating approach for the last 120 thousand years. *Clim. Past Discuss.* **8**, 6011–6049 (2012).
- Veres, D. *et al.* The Antarctic ice core chronology (AICC2012): an optimized multi-parameter and multi-site dating approach for the last 120 thousand years. *Clim. Past* **9**, 1733–1748 (2013).
- Grobe, H. A simple method for the determination of ice-rafted debris in sediment cores. *Polarforschung* **57**, 123–126 (1987).
- Roy, M., van de Flierdt, T., Hemming, S. R. & Goldstein, S. L.  $^{40}\text{Ar}/^{39}\text{Ar}$  ages of hornblende grains and bulk Sm/Nd isotopes of circum-Antarctic glacio-marine sediments: implications for sediment provenance in the southern ocean. *Chem. Geol.* **244**, 507–519 (2007).
- Pierce, E. L. *et al.* Characterizing the sediment provenance of East Antarctica’s weak underbelly: the Aurora and Wilkes sub-glacial basins. *Paleoceanography* **26**, PA4217 (2011).
- Galton-Fenzi, B. K., Hunter, J. R., Coleman, R., Marsland, S. J. & Warner, R. C. Modeling the basal melting and marine ice accretion of the Amery Ice Shelf. *J. Geophys. Res.* **117**, C09031 (2012).

56. Anderson, J. B. & Andrews, J. T. Radiocarbon constraints on ice sheet advance and retreat in the Weddell Sea, Antarctica. *Geology* **27**, 179–182 (1999).
57. Schodlok, M. P., Hellmer, H. H., Rohardt, G. & Fahrbach, E. Weddell Sea ice-berg drift: five years of observations. *J. Geophys. Res.* **111**, C06018 (2006).
58. Pudsey, C. J. & Howe, J. A. Quaternary history of the Antarctic Circumpolar Current: evidence from the Scotia Sea. *Mar. Geol.* **148**, 83–112 (1998).
59. Jacka, T. H. & Giles, A. B. Antarctic iceberg distribution and dissolution from ship-based observations. *J. Glaciol.* **53**, 341–356 (2007).
60. Williams, T. *et al.* Evidence for iceberg armadas from East Antarctica in the Southern Ocean during the late Miocene and early Pliocene. *Earth Planet. Sci. Lett.* **290**, 351–361 (2010).
61. Toggweiler, J. R. & Russell, J. Ocean circulation in a warming climate. *Nature* **451**, 286–288 (2008).
62. Toggweiler, J. R. & Lea, D. W. Temperature differences between the hemispheres and ice age climate variability. *Paleoceanography* **25**, PA2212 (2010).
63. Smith, J. A., Hillenbrand, C.-D., Pudsey, C. J., Allen, C. S. & Graham, A. G. C. The presence of polynyas in the Weddell Sea during the Last Glacial Period with implications for the reconstruction of sea-ice limits and ice sheet history. *Earth Planet. Sci. Lett.* **296**, 287–298 (2010).
64. Nielsen, S. H. H., Hodell, D. A., Kamenov, G., Guilderson, T. & Perfit, M. R. Origin and significance of ice-rafted detritus in the Atlantic sector of the Southern Ocean. *Geochem. Geophys. Geosyst.* **8**, 1–23 (2007).
65. Nakada, M. *et al.* Late Pleistocene and Holocene melting history of the Antarctic ice sheet derived from sea-level variations. *Mar. Geol.* **167**, 85–103 (2000).
66. White, D. A., Fink, D. & Gore, D. B. Cosmogenic nuclide evidence for enhanced sensitivity of an East Antarctic ice stream to change during the last deglaciation. *Geology* **39**, 23–26 (2011).
67. Kirshner, A. E. *et al.* Post-LGM deglaciation in Pine Island Bay, West Antarctica. *Quat. Sci. Rev.* **38**, 11–26 (2012).
68. Hall, B. L. & Denton, G. H. Radiocarbon chronology of Ross Sea drift, Eastern Taylor Valley, Antarctica: evidence for a grounded ice sheet in the Ross Sea at the Last Glacial Maximum. *Geogr. Ann.* **82**, 305–336 (2000).
69. Price, S. F., Conway, H. & Waddington, E. D. Evidence for late Pleistocene thinning of Siple Dome, West Antarctica. *J. Geophys. Res.* **112**, F03021 (2007).
70. Simms, A. R., Milliken, K. T., Anderson, J. B. & Wellner, J. S. The marine record of deglaciation of the South Shetland Islands, Antarctica since the Last Glacial Maximum. *Quat. Sci. Rev.* **30**, 1583–1601 (2011).
71. Leventer, A. *et al.* Marine sediment record from the East Antarctic margin reveals dynamics of ice sheet recession. *Geol. Soc. Am. Today* **16**, 4–10 (2006).
72. McKay, R. M. *et al.* Retreat history of the Ross Ice Sheet (Shelf) since the Last Glacial Maximum from deep-basin sediment cores around Ross Island. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **260**, 245–261 (2008).
73. Menviel, L. & Joos, F. Toward explaining the Holocene carbon dioxide and carbon isotope records: results from transient ocean carbon cycle-climate simulations. *Paleoceanography* **27**, PA1207 (2012).
74. Ritz, S. P., Stocker, T. F. & Joos, F. A coupled dynamical ocean–energy balance atmosphere model for paleoclimate studies. *J. Clim.* **24**, 349–375 (2011).
75. Menviel, L., Timmermann, A., Mouchet, A. & Timm, O. Climate and marine carbon cycle response to changes in the strength of the Southern Hemispheric westerlies. *Paleoceanography* **23**, PA4201 (2008).
76. Timmermann, A., Timm, O., Stott, L. & Menviel, L. The roles of CO<sub>2</sub> and orbital forcing in driving Southern Hemispheric temperature variations during the last 21 000 Yr. *J. Clim.* **22**, 1626–1640 (2009).
77. Timm, O., Timmermann, A., Abe-Ouchi, A., Saito, F. & Segawa, T. On the definition of seasons in paleoclimate simulations with orbital forcing. *Paleoceanography* **23**, PA2221 (2008).
78. Knorr, G., Butzin, M., Micheels, A. & Lohmann, G. A warm Miocene climate at low atmospheric CO<sub>2</sub> levels. *Geophys. Res. Lett.* **38**, L20701 (2011).
79. Zhang, X., Lohmann, G., Knorr, G. & Xu, X. Different ocean states and transient characteristics in Last Glacial Maximum simulations and implications for deglaciation. *Clim. Past* **9**, 2319–2333 (2013).
80. Wei, W., Lohmann, G. & Dima, M. Distinct modes of internal variability in the global meridional overturning circulation associated with the Southern Hemisphere westerly winds. *J. Phys. Oceanogr.* **42**, 785–801 (2012).
81. Laurantou, A. *et al.* Constraint of the CO<sub>2</sub> rise by new atmospheric carbon isotopic measurements during the last deglaciation. *Glob. Biogeochem. Cycles* **24**, GB2015 (2010).



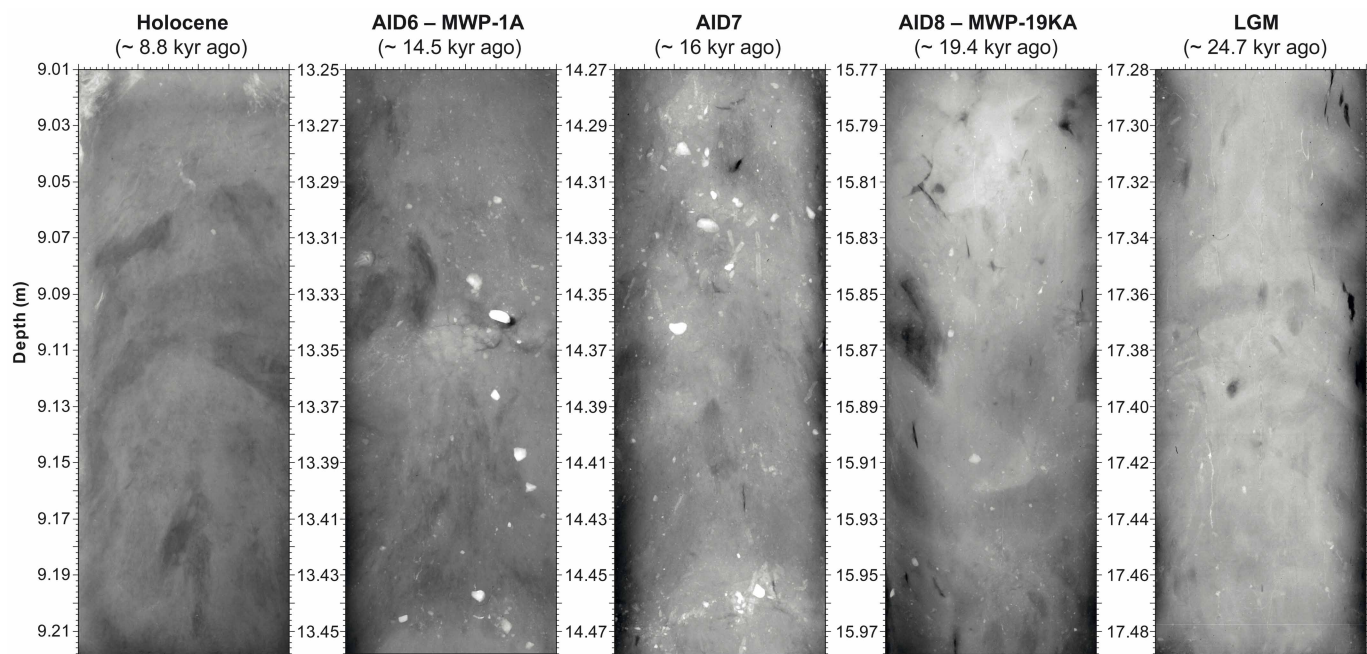


**Extended Data Figure 1 | Deglacial dust chronology.** Five common tie points (TP1 to TP5, indicated by green vertical bands) depict consistent changes in slope and reproducible lows and highs between magnetic susceptibility (a, b), Fe (c) and Ca (d) records of deep-sea sites MD07-31333 and MD07-313412, and the non-sea-salt Ca record (e) of the EDML ice core<sup>27</sup>.



**Extended Data Figure 2 | Uncertainty estimates for AIDs.** Conservative error estimates ( $2\sigma$ ) rely on bootstrapping of different age models and projecting them on to AIDs. **a**, Errors of the MD age model<sup>12</sup> based on tie point correlation only. Black dots depict the centre of the AID and its absolute uncertainty range. Black error bars at the boxes mark the relative uncertainty with respect to the centre. Grey error bars show the absolute uncertainty of the beginning and end of each AID. **b**, Errors including the EDML1 (ref. 47), and

EDC3 (ref. 46) uncertainties. **c**, Relative duration of AIDs and related uncertainties. **d**, Error propagation of the three different age scales through the last deglaciation. IU is interpolation uncertainty. Note that uncertainties are highly correlated for nearby ages. Accounting for this correlation, the duration of each AID as well as the time between two AIDs is significantly more accurate than its absolute age uncertainty.



**Extended Data Figure 3 | X-radiograph images from Scotia Sea Site MD07-3134.** IBRD (bright dropstones) are embedded in a matrix-supported diatomaceous mud. Low IBRD contents are documented for the Last Glacial Maximum (LGM, ~24.7 kyr ago) and the Holocene (~8.8 kyr ago), whereas

higher numbers indicate enhanced iceberg routing through Iceberg Alley during three distinct deglaciation phases (centre panels): AID8 (MWP-19KA), AID7 and AID6 (MWP-1A).



Extended Data Table 1 | Uncertainty estimates

AID No.	Top (kyr)	Error ( $2\sigma$ )	Centre (kyr)	Error ( $2\sigma$ )	Bottom (kyr)	Error ( $2\sigma$ )
AID1	8.99	(0.32, 0.38)	9.63	(0.29, 0.38)	10.28	(0.27, 0.38)
AID2	11.00	(0.26, 0.39)	11.23	(0.25, 0.39)	11.46	(0.25, 0.39)
AID3	11.78	(0.24, 0.40)	11.90	(0.24, 0.41)	12.03	(0.24, 0.42)
AID4	12.71	(0.22, 0.42)	12.88	(0.22, 0.42)	13.04	(0.22, 0.42)
AID5	13.28	(0.21, 0.43)	13.46	(0.21, 0.43)	13.65	(0.20, 0.44)
AID6	13.94	(0.19, 0.44)	14.40	(0.17, 0.50)	14.86	(0.16, 0.55)
AID7	15.75	(0.13, 0.67)	16.33	(0.13, 0.76)	16.91	(0.13, 0.85)
AID8	19.00	(0.07, 1.03)	19.54	(0.08, 1.04)	20.08	(0.09, 1.05)

Top, bottom, and centre of deglacial AID1 to AID8, as well as absolute dating uncertainties in terms of the  $2\sigma$  error (parenthesis). The first uncertainty relates to the MD07 chronology only<sup>12</sup>, the one below to the overall uncertainty including the EDML1 (ref. 47) and EDC3 (ref. 46) chronology.

# Increasing CO<sub>2</sub> threatens human nutrition

Samuel S. Myers<sup>1,2</sup>, Antonella Zanutti<sup>1</sup>, Itai Kloog<sup>3</sup>, Peter Huybers<sup>4</sup>, Andrew D. B. Leakey<sup>5</sup>, Arnold J. Bloom<sup>6</sup>, Eli Carlisle<sup>6</sup>, Lee H. Dietterich<sup>7</sup>, Glenn Fitzgerald<sup>8</sup>, Toshihiro Hasegawa<sup>9</sup>, N. Michele Holbrook<sup>10</sup>, Randall L. Nelson<sup>11</sup>, Michael J. Ottman<sup>12</sup>, Victor Raboy<sup>13</sup>, Hidemitsu Sakai<sup>9</sup>, Karla A. Sartor<sup>14</sup>, Joel Schwartz<sup>1</sup>, Saman Seneweera<sup>15</sup>, Michael Tausz<sup>16</sup> & Yasuhiro Usui<sup>9</sup>

**Dietary deficiencies of zinc and iron are a substantial global public health problem. An estimated two billion people suffer these deficiencies<sup>1</sup>, causing a loss of 63 million life-years annually<sup>2,3</sup>. Most of these people depend on C<sub>3</sub> grains and legumes as their primary dietary source of zinc and iron. Here we report that C<sub>3</sub> grains and legumes have lower concentrations of zinc and iron when grown under field conditions at the elevated atmospheric CO<sub>2</sub> concentration predicted for the middle of this century. C<sub>3</sub> crops other than legumes also have lower concentrations of protein, whereas C<sub>4</sub> crops seem to be less affected. Differences between cultivars of a single crop suggest that breeding for decreased sensitivity to atmospheric CO<sub>2</sub> concentration could partly address these new challenges to global health.**

In the 1990s, several investigators found that elevated atmospheric CO<sub>2</sub> concentration (hereafter abbreviated to [CO<sub>2</sub>]) decreased the concentrations of zinc, iron and protein in grains of wheat<sup>4–7</sup>, barley<sup>5</sup> and rice<sup>8</sup> grown in controlled-environment chambers. However, subsequent studies failed to replicate these results when plants were grown in open-top chambers and free-air CO<sub>2</sub> enrichment (FACE) experiments. A previous study<sup>9</sup> found no effect of [CO<sub>2</sub>] on the concentrations of zinc or iron in rice grains grown under FACE and suggested that the earlier findings had been influenced by ‘pot effects’, by which a small rooting volume led to nutrient dilution at the root–soil interface. Of the more recent studies<sup>10–13</sup>, most have indicated lower elemental concentrations in soybeans<sup>10</sup>, sorghum<sup>10</sup>, potatoes<sup>11</sup>, wheat<sup>12</sup> or barley<sup>13</sup> grown at elevated [CO<sub>2</sub>], but with the exception of iron in one study on wheat<sup>12</sup>, these results were statistically insignificant, perhaps because of small sample sizes.

Small sample sizes have limited the statistical power of individual studies of many aspects of plant responses to elevated [CO<sub>2</sub>], and meta-analyses involving larger samples of genotypes, environmental conditions and experimental locations have been important in resolving which elements of plant function respond reliably to altered [CO<sub>2</sub>]<sup>14,15</sup>. A recent meta-analysis of published data concluded that only sulphur is decreased in grains grown at elevated [CO<sub>2</sub>]<sup>16</sup>.

Here we report findings from a meta-analysis of newly acquired data from 143 comparisons of the edible portions of crops grown at ambient and elevated [CO<sub>2</sub>] from seven different FACE experimental locations in Japan, Australia and the United States involving six food crops (see Table 1). We tested the nutrient concentrations of the edible portions of rice (*Oryza sativa*, 18 cultivars), wheat (*Triticum aestivum*, 8 cultivars), maize (*Zea mays*, 2 cultivars), soybeans (*Glycine max*, 7 cultivars), field peas (*Pisum sativum*, 5 cultivars) and sorghum (*Sorghum bicolor*, 1 cultivar). In all, forty-one genotypes were tested over one to six growing seasons at ambient and elevated [CO<sub>2</sub>], where the latter was in the range 546–586 p.p.m. across all seven study sites. Collectively, these

experiments contribute more than tenfold more data regarding both the zinc and iron content of the edible portions of crops grown under FACE conditions than is currently available in the literature. Consistent with earlier meta-analyses of other aspects of plant function under FACE conditions<sup>14,15</sup>, we considered the response comparisons observed from different species, cultivars and stress treatments and from different years to be independent. The natural logarithm of the mean response ratio ( $r$  = response in elevated [CO<sub>2</sub>]/response in ambient [CO<sub>2</sub>]) was used as the metric for all analyses. Meta-analysis was used to estimate the overall effect of elevated [CO<sub>2</sub>] on the concentration of each nutrient in a particular crop and to determine the significance of this effect (see Methods).

We found that elevated [CO<sub>2</sub>] was associated with significant decreases in the concentrations of zinc and iron in all C<sub>3</sub> grasses and legumes (Fig. 1 and Extended Data Table 1). For example, wheat grains grown at elevated [CO<sub>2</sub>] had 9.3% lower zinc (95% confidence interval (CI) –12.7% to –5.9%) and 5.1% lower iron (95% CI –6.5% to –3.7%) than those grown at ambient [CO<sub>2</sub>]. We also found that elevated [CO<sub>2</sub>] was associated with lower protein content in C<sub>3</sub> grasses, with a 6.3% decrease (95% CI –7.5% to –5.2%) in wheat grains and a 7.8% decrease (95% CI –8.9% to –6.8%) in rice grains. Elevated [CO<sub>2</sub>] was associated with a small decrease in protein in field peas, and there was no significant effect in soybeans or C<sub>4</sub> crops (Fig. 1 and Extended Data Table 1).

In addition to our own observations, we obtained data from 10 of 11 previously published studies investigating nutrient changes in the edible portion of food crops (Extended Data Table 6) and combined these data with our own observations in a larger meta-analysis. Analysis of our results combined with previously published FACE data (Extended Data Table 2), or combined with previously published data from both FACE and chamber experiments (Extended Data Table 3), was consistent with the results obtained using only our new data. Combining our data with previously published data did not alter the significance or substantially alter the effect size of the nutrient changes for any crop or any nutrient.

In addition to nutrient concentrations, we also measured phytate, a phosphate storage molecule present in most plants that inhibits the absorption of dietary zinc in the human gut<sup>17</sup>. We had no a priori reason to assume that phytate concentrations would change in response to rising [CO<sub>2</sub>]. However, formulae for calculating absorbed, or bioavailable, zinc depend on both the amount of dietary zinc and the amount of dietary phytate consumed<sup>17</sup>, making it important to interpret changes in zinc concentration in the context of possible changes in phytate. Phytate content decreased significantly at elevated [CO<sub>2</sub>] only in wheat ( $P < 0.01$ ). This decrease might offset some of the declines in zinc for this particular crop, although the decrease was slightly less than half of the decrease in zinc. For other crops examined, however, the lack of a concurrent decrease in phytate may further exacerbate problems of zinc deficiency.

<sup>1</sup>Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, 02215, USA. <sup>2</sup>Harvard University Center for the Environment, Cambridge, Massachusetts 02138, USA. <sup>3</sup>The Department of Geography and Environmental Development, Ben-Gurion University of the Negev, PO Box 653, Beer Sheva, Israel. <sup>4</sup>Department of Earth and Planetary Science, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>5</sup>Department of Plant Biology and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. <sup>6</sup>Department of Plant Sciences, University of California at Davis, Davis, California 95616, USA. <sup>7</sup>University of Pennsylvania, Department of Biology, Philadelphia, Pennsylvania 19104, USA. <sup>8</sup>Department of Environment and Primary Industries, Horsham, Victoria 3001, Australia. <sup>9</sup>National Institute for Agro-Environmental Sciences, Tsukuba, Ibaraki, 305-8604, Japan. <sup>10</sup>Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>11</sup>United States Department of Agriculture Agricultural Research Service, Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, Department of Crop Sciences, University of Illinois, Urbana, Illinois 61801, USA. <sup>12</sup>School of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA. <sup>13</sup>United States Department of Agriculture Agricultural Research Service, Aberdeen, Idaho 83210, USA. <sup>14</sup>The Nature Conservancy, Santa Fe, New Mexico 87544, USA. <sup>15</sup>Department of Agriculture and Food Systems, Melbourne School of Land and Environment, The University of Melbourne, Creswick, Victoria 3363, Australia. <sup>16</sup>Department of Forest and Ecosystem Science, Melbourne School of Land and Environment, The University of Melbourne, Creswick, Victoria 3363, Australia.

**Table 1 | Characteristics of agricultural experiments**

Crops	Country	Treatments used	Years grown	Number of replicates	Number of cultivars	CO <sub>2</sub> ambient/elevated (p.p.m.)
Wheat						
Site 1	Australia	2 water levels, 2 nitrogen treatments, 2 sowing times	2007–2010	4	8	382/546–550
Site 2	Australia	1 water level, 1 nitrogen treatment, 2 sowing times	2007–2009	4	1	382/546–550
Field peas	Australia	2 water levels	2010	4	5	382/546–550
Rice						
Site 1	Japan	1 nitrogen treatment, 2 warming treatments	2007–2008	3	3	376–379/570–576
Site 2	Japan	3 nitrogen treatments, 2 warming treatments	2010	4	18	386/584
Maize	United States	2 nitrogen treatments	2008	4	2	385/550
Soybeans	United States	1 treatment	2001, 2002, 2004, 2006–2008	4	7	372–385/550
Sorghum	United States	2 water levels	1998–1999	4	1	363–373/556–579

<sup>a</sup>Number of replicates refers to the number of identical cultivars grown under identical conditions in the same year and location but in separate FACE rings.

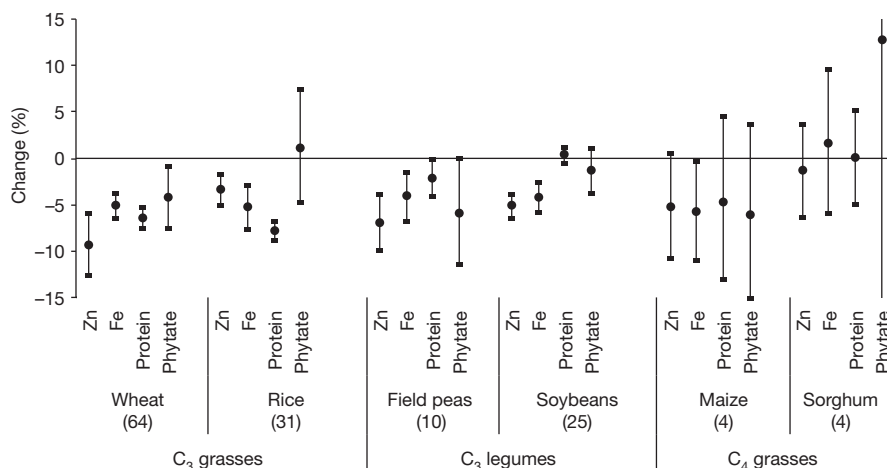
The global [CO<sub>2</sub>] in the atmosphere is expected to reach 550 p.p.m. in the next 40–60 years, even if further actions are taken to decrease emissions<sup>18</sup>. At these concentrations, we find that the edible portions of many of the key crops for human nutrition have decreased nutritional value when compared with the same plants grown under identical conditions but at the present ambient [CO<sub>2</sub>]. Analysis of the United Nations' Food and Agriculture Organization food balance sheets reveals that in 2010 roughly 2.3 billion people were living in countries whose populations received at least 60% of their dietary zinc and/or iron from C<sub>3</sub> grains and legumes, and 1.9 billion lived in countries that received at least 70% of one or both of these nutrients from these crops (Extended Data Table 5). Reductions in the zinc and iron content of the edible portion of these food crops will increase the risk of zinc and iron deficiencies across these populations and will add to the already considerable burden of disease associated with them.

The implications of decreased protein concentrations in non-leguminous C<sub>3</sub> crops are less clear. From a study of adult men and women in the United States, there is strong evidence that the substitution of dietary carbohydrate for dietary protein increased the risk of hypertension, lipid disorders, and 10-year coronary heart disease risk<sup>19</sup>. For the developing world, minimum protein requirements for different demographic groups are an area of active research and debate<sup>20</sup>. For countries such as India, however, in which up to one-third of the rural population is thought to be at risk of not meeting protein requirements<sup>21</sup> and in which most

protein comes in the form of C<sub>3</sub> grains<sup>21</sup>, decreased protein content in non-leguminous C<sub>3</sub> crops may have serious consequences for public health.

Whereas zinc and iron were significantly decreased in all C<sub>3</sub> crops tested, only iron in maize was observed to decrease among the C<sub>4</sub> crops. No changes were found in sorghum. That zinc and iron declines were notable in C<sub>3</sub> crops but less so in C<sub>4</sub> crops is consistent with differences in physiology. C<sub>4</sub> crops concentrate CO<sub>2</sub> internally, which results in photosynthesis being CO<sub>2</sub>-saturated even under ambient [CO<sub>2</sub>] conditions, leading to no stimulation of photosynthetic carbon assimilation at elevated [CO<sub>2</sub>] levels under mesic growing conditions<sup>22</sup>. Our finding that protein content was less affected in legumes than in other C<sub>3</sub> crops is also physiologically consistent with the general ability of leguminous crops to match the stimulation of photosynthetic carbon gain at elevated [CO<sub>2</sub>] with greater nitrogen fixation, to maintain tissue carbon:nitrogen (C:N) ratios<sup>23</sup>. In contrast, most temperate non-legume C<sub>3</sub> crops are generally unable to extract and assimilate sufficient nitrogen from soils to maintain tissue C:N ratios<sup>24,25</sup>.

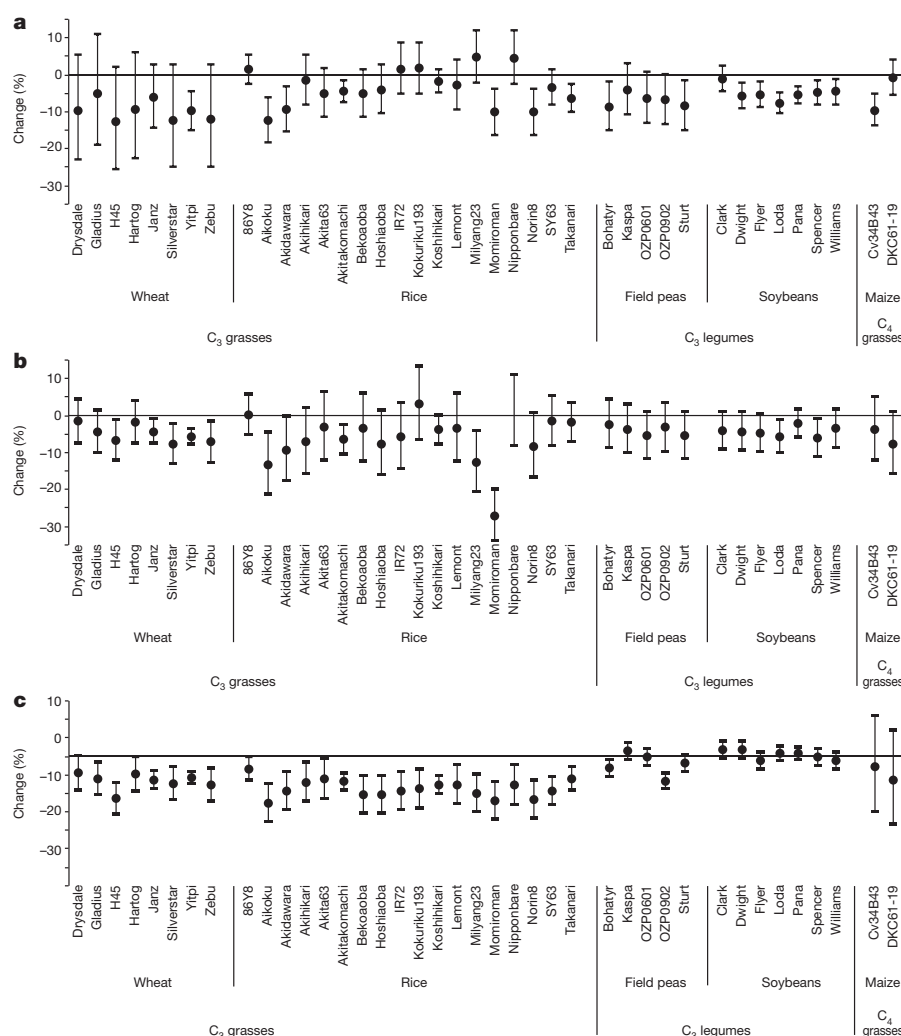
Little is known about the mechanism(s) responsible for the decline in nutrient concentrations associated with elevated [CO<sub>2</sub>]. Some authors have proposed 'carbohydrate dilution', by which CO<sub>2</sub>-stimulated carbohydrate production by plants dilutes the rest of the grain components<sup>26</sup>. To test this hypothesis, we measured concentrations of additional elements for all crops except wheat (Extended Data Table 4). Our findings were inconsistent with carbohydrate dilution operating alone. If only



**Figure 1 | Percentage change in nutrients at elevated [CO<sub>2</sub>] relative to ambient [CO<sub>2</sub>].** Numbers in parentheses refer to the number of comparisons in which replicates of a particular cultivar grown at a specific site under one set of growing conditions in one year at elevated [CO<sub>2</sub>] have been pooled and for which mean nutrient values for these replicates are compared with mean values

for identical cultivars under identical growing conditions except grown at ambient [CO<sub>2</sub>]. In most instances, data from four replicates were pooled for each value, meaning that eight experiments were combined for each comparison (see Table 1 for details of experiments). Error bars represent 95% confidence intervals of the estimates.





**Figure 2 | Percentage change (with 95% confidence intervals) in nutrients at elevated  $[\text{CO}_2]$  relative to ambient  $[\text{CO}_2]$ , by cultivar. a, Zinc; b, iron; c, protein.**

passive dilution of nutrients were occurring, we would have expected to see very similar changes in the concentration of each nutrient tested for a given crop. In contrast, we found that elemental changes in the individual crops are distinct from each other. For example, in rice grains (Extended Data Table 4) the decrease in zinc concentrations associated with elevated  $[\text{CO}_2]$  was significantly different from the decreases in the concentrations of copper ( $P \leq 0.001$ ), calcium ( $P \leq 0.001$ ), boron ( $P \leq 0.001$ ) and phosphate ( $P = 0.010$ ). This heterogeneous response was also observed in recent analyses reviewing possible mechanisms for nutrient changes in both edible and non-edible plant tissues grown at elevated  $[\text{CO}_2]$ <sup>27</sup>. It also seems that the mechanism(s) causing these changes operate distinctly in different species. In one instance, for example, we found boron to be significantly decreased in soybeans ( $P \leq 0.001$ ), whereas it was significantly elevated in rice grains ( $P \leq 0.001$ ). Although these differences may, in part, have derived from different environmental conditions, they suggest that the mechanism is more complex than carbohydrate dilution alone. Of all the elements, changes in nitrogen content at elevated  $[\text{CO}_2]$  have been the most studied, and inhibition of photorespiration and malate production<sup>24</sup>, carbohydrate dilution<sup>26</sup>, slower uptake of nitrogen in roots<sup>25</sup> and decreased transpiration-driven mass flow of nitrogen<sup>7</sup> may all be significant.

We also examined the effects of elevated  $[\text{CO}_2]$  on zinc, iron and protein content as a function of cultivar when data were available (Fig. 2). Whereas most crops showed negligible differences across cultivars, concentrations of zinc and iron across rice cultivars varied substantially ( $P = 0.04$  and  $P = 0.03$ , respectively; Fig. 2a, b). Such differences between

cultivars suggest a basis for breeding rice cultivars whose micronutrient levels are less vulnerable to increasing  $[\text{CO}_2]$ . Similar effects may occur in other crops, given that the statistical power of many of our other inter-cultivar tests was limited by sample size. We note, however, that such breeding programmes will not be a panacea for many reasons including the affordability of improved seeds and the numerous criteria used by farmers in making planting decisions that include taste, tradition, marketability, growing requirements and yield. In addition, as has been noted previously, there are likely to be trade-offs with respect to yield and other performance characteristics when breeding for increased zinc and iron content<sup>28</sup>.

The public health implications of global climate change are difficult to predict, and we expect many surprises. The finding that raising atmospheric  $[\text{CO}_2]$  lowers the nutritional value of  $\text{C}_3$  food crops is one such surprise that we can now better predict and prepare for. In addition to efforts to limit increases in  $[\text{CO}_2]$ , it may be important to develop breeding programmes designed to decrease the vulnerability of key crops to these changes. Nutritional analysis of which human populations are most vulnerable to decreased dietary availability of zinc, iron and protein from  $\text{C}_3$  crops could help to target response efforts, including breeding decreased sensitivity to elevated  $[\text{CO}_2]$ , biofortification, and supplementation.

## METHODS SUMMARY

We examined the response of nutrient levels to elevated atmospheric  $[\text{CO}_2]$  for the edible portions of rice (*Oryza sativa*, 18 cultivars), wheat (*Triticum aestivum*, 8 cultivars), maize (*Zea mays*, 2 cultivars), soybeans (*Glycine max*, 7 cultivars), field

peas (*Pisum sativum*, 5 cultivars) and sorghum (*Sorghum bicolor*, 1 cultivar). The six crops were grown under FACE conditions; in all six experiments the elevated [CO<sub>2</sub>] was in the range 546–586 p.p.m.

In accordance with methods described previously<sup>14,15</sup>, the natural logarithm of the response ratio ( $r$  = response in elevated [CO<sub>2</sub>]/response in ambient [CO<sub>2</sub>]) was used as the metric for analyses and is reported as the mean percentage change ( $100 \times (r - 1)$ ) at elevated [CO<sub>2</sub>]. Consistent with these earlier analyses of multiple species grown under FACE conditions, the responses of different species, cultivars and stress treatments and from different years of the FACE experiments were considered to be independent and suited to meta-analytic analysis<sup>14</sup>.

The meta-analysis was designed to estimate the effect of elevated [CO<sub>2</sub>] on the concentration of each nutrient in a particular crop and to determine the significance of this effect relative to a null hypothesis of no change. All tests were conducted as two-sided; that is, not specifying which direction the nutrient concentrations were expected to change under elevated [CO<sub>2</sub>]. Meta-analysis was conducted with a linear mixed model.

Parameter estimates were obtained by the restricted maximum-likelihood method, a standard approach for analysing repeated measurements<sup>29</sup> that, in our case, were of nutrient concentrations at the time of harvest. Results for all analyses are reported as the best estimate of percentage changes in the concentration of nutrients along with the 95% confidence intervals associated with each estimate. Two-tailed *P* values are also reported.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 November 2013; accepted 24 February 2014.

Published online 7 May 2014.

- Tulchinsky, T. H. Micronutrient deficiency conditions: global health issues. *Public Health Rev.* **32**, 243–255 (2010).
- Caulfield, L. E. & Black, R. E. in *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attribution to Selected Major Risk Factors* (eds Ezzati, M., Lopez, A. D., Rodgers, A. & Murray, C. J. L.) Vol. 1, Ch. 5 (World Health Organization, 2004).
- Stoltzfus, R. J., Mullany, L. & Black, R. E. in *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attribution to Selected Major Risk Factors* (eds Ezzati, M., Lopez, A. D., Rodgers, A. & Murray, C. J. L.) Vol. 1, Ch. 3 (World Health Organization, 2004).
- De la Puente, L. S., Pérez, P. P., Martínez-Carrasco, R., Morcuende, R. M. & Del Molino, I. M. M. Action of elevated CO<sub>2</sub> and high temperatures on the mineral chemical composition of two varieties of wheat. *Agrochimica* **44**, 221–230 (2000).
- Manderscheid, R., Bender, J., Jäger, H. J. & Weigel, H. J. Effects of season long CO<sub>2</sub> enrichment on cereals. II. Nutrient concentrations and grain quality. *Agric. Ecosyst. Environ.* **54**, 175–185 (1995).
- Fangmeier, A., Grüters, U., Högy, P., Vermehren, B. & Jäger, H.-J. Effects of elevated CO<sub>2</sub>, nitrogen supply and tropospheric ozone on spring wheat. II. Nutrients (N, P, K, S, Ca, Mg, Fe, Mn, Zn). *Environ. Pollut.* **96**, 43–59 (1997).
- Pleijel, H. et al. Effects of elevated carbon dioxide, ozone and water availability on spring wheat growth and yield. *Physiol. Plant.* **108**, 61–70 (2000).
- Seneweera, S. P. & Conroy, J. P. Growth, grain yield and quality of rice (*Oryza sativa* L.) in response to elevated CO<sub>2</sub> and phosphorus nutrition. *Soil Sci. Plant Nutr.* **43**, 1131–1136 (1997).
- Lieffering, M., Kim, H.-Y., Kobayashi, K. & Okada, M. The impact of elevated CO<sub>2</sub> on the elemental concentrations of field-grown rice grains. *Field Crops Res.* **88**, 279–286 (2004).
- Prior, S. A., Runion, G. B., Rogers, H. H. & Torbert, H. A. Effects of atmospheric CO<sub>2</sub> enrichment on crop nutrient dynamics under no-till conditions. *J. Plant Nutr.* **31**, 758–773 (2008).
- Högy, P. & Fangmeier, A. Atmospheric CO<sub>2</sub> enrichment affects potatoes. 2. Tuber quality traits. *Eur. J. Agron.* **30**, 85–94 (2009).
- Högy, P. et al. Effects of elevated CO<sub>2</sub> on grain yield and quality of wheat: results from a 3-year free-air CO<sub>2</sub> enrichment experiment. *Plant Biol.* **11**, 60–69 (2009).
- Erbs, M. et al. Effects of free-air CO<sub>2</sub> enrichment and nitrogen supply on grain quality parameters and elemental composition of wheat and barley grown in a crop rotation. *Agric. Ecosyst. Environ.* **136**, 59–68 (2010).
- Ainsworth, E. A. & Long, S. P. What have we learned from 15 years of free-air CO<sub>2</sub> enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO<sub>2</sub>. *New Phytol.* **165**, 351–372 (2005).
- Curtis, P. S. & Wang, X. A meta-analysis of elevated CO<sub>2</sub> effects on woody plant mass, form, and physiology. *Oecologia* **113**, 299–313 (1998).
- Duval, B. D., Blankinship, J. C., Dijkstra, P. & Hungate, B. A. CO<sub>2</sub> effects on plant nutrient concentration depend on plant functional group and available nitrogen: a meta-analysis. *Plant Ecol.* **213**, 505–521 (2012).
- Miller, L. V., Krebs, N. F. & Hambidge, M. K. A mathematical model of zinc absorption in humans as a function of dietary zinc and phytate. *J. Nutr.* **137**, 135–141 (2007).
- Fisher, B. S. et al. in *Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Metz, B. et al.) 169–250 (Cambridge Univ. Press, 2007).
- Appel, L. J. et al. Effects of protein, monounsaturated fat, and carbohydrate intake on blood pressure and serum lipids: results of the OmniHeart randomized trial. *J. Am. Med. Assoc.* **294**, 2455–2464 (2005).
- Millward, D. Joe. Identifying recommended dietary allowances for protein and amino acids: a critique of the 2007 WHO/FAO/UNU report. *Br. J. Nutr.* **108**, S3–S21 (2012).
- Swaminathan, S., Vaz, M. & Kurpad, A. V. Protein intakes in India. *Br. J. Nutr.* **108**, S50–S58 (2012).
- Leakey, A. Rising atmospheric carbon dioxide concentration and the future of C<sub>4</sub> crops for food and fuel. *Proc. R. Soc. Lond. B* **276**, 2333–2343 (2009).
- Rogers, A., Ainsworth, E. A. & Leakey, A. D. Will elevated carbon dioxide concentration amplify the benefits of nitrogen fixation in legumes? *Plant Physiol.* **151**, 1009–1016 (2009).
- Bloom, A. J. et al. CO<sub>2</sub> enrichment inhibits shoot nitrate assimilation in C<sub>3</sub> but not C<sub>4</sub> plants and slows growth under nitrate in C<sub>3</sub> plants. *Ecology* **93**, 355–367 (2012).
- Leakey, A. D. et al. Elevated CO<sub>2</sub> effects on plant carbon, nitrogen, and water relations: six important lessons from FACE. *J. Exp. Bot.* **60**, 2859–2876 (2009).
- Gifford, R., Barrett, D. & Lutze, J. The effects of elevated [CO<sub>2</sub>] on the C:N and C:P mass ratios of plant tissues. *Plant Soil* **224**, 1–14, 10.1023/A:1004790612630 (2000).
- McGrath, J. M. & Lobell, D. B. Reduction of transpiration and altered nutrient allocation contribute to nutrient decline of crops grown in elevated CO<sub>2</sub> concentrations. *Plant Cell Environ.* **36**, 697–705, 10.1111/pce.12007 (2013).
- Monasterio, I. & Graham, R. D. Breeding for trace minerals in wheat. *Food Nutr. Bull.* **21**, 392–396 (2000).
- Searle, S. R., Casella, G. & McCulloch, C. E. *Variance Components* (Wiley, 1992).

**Acknowledgements** We thank L. S. De la Puente, M. Erbs, A. Fangmeier, P. Högy, M. Lieffering, R. Manderscheid, H. Pleijel and S. Prior for sharing data from their groups with us; H. Nakamura, T. Tokida, C. Zhu and S. Yoshinaga for contributions to the rice FACE project; and M. Hambidge, W. Willett, D. Schrag, K. Brown, R. Wessells, N. Fernando, J. Pearson and B. Kimball for reviews of earlier drafts or conceptual contributions to this project. V.R. thanks A. L. Harvey for her efforts in producing the phytate data included here. The National Agriculture and Food Research Organization (Japan) provided the grain samples of some rice cultivars. We thank the following for financial support of this work: the Bill & Melinda Gates Foundation; the Winslow Foundation; the Commonwealth Department of Agriculture (Australia), the International Plant Nutrition Institute, (Australia), the Grains Research and Development Corporation (Australia), the Ministry of Agriculture, Forestry and Fisheries (Japan); the National Science Foundation (NSF IOS-08-18435); USDA NIFA 2008-35100-044459; research at SoyFACE was supported by the US Department of Agriculture Agricultural Research Service; Illinois Council for Food and Agricultural Research (CFAR); Department of Energy's Office of Science (BER) Midwestern Regional Center of the National Institute for Climatic Change Research at Michigan Technological University, under Award Number DEFC02-06ER64158; and the National Research Initiative of Agriculture and Food Research Initiative Competitive Grants Program Grant no. 2010-65114-20343 from the USDA National Institute of Food and Agriculture. Early stages of this work received support from Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health Award 8UL1TR000170-05).

**Author Contributions** S.S.M. conceived the overall project and drafted the manuscript. A.Z., I.K., J.S. and P.H. performed statistical analyses. P.H. and A.D.B.L. provided substantial input into methods descriptions. A.J.B., E.C. and V.R. analysed grain samples for nutrient content. G.F., T.H., A.D.B.L., R.L.N., M.J.O., H.S., S.S., M.T. and Y.U. conducted FACE experiments and supplied grain for analysis. N.M.H. and P.H. assisted with elements of experimental design. K.A.S. and L.H.D. assisted with data collection and analysis. All authors contributed to manuscript preparation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.S.M. ([smymers@hsph.harvard.edu](mailto:smymers@hsph.harvard.edu)).

## METHODS

We examined the response of nutrient levels to elevated atmospheric [CO<sub>2</sub>] for the edible portions of rice (*Oryza sativa*, 18 cultivars), wheat (*Triticum aestivum*, 8 cultivars), maize (*Zea mays*, 2 cultivars), soybeans (*Glycine max*, 7 cultivars), field peas (*Pisum sativum*, 5 cultivars) and sorghum (*Sorghum bicolor*, 1 cultivar). The six crops were grown under FACE conditions; in all six experiments, the elevated [CO<sub>2</sub>] was in the range 546–586 p.p.m. (see the Agricultural Methods section below for details associated with individual trials).

**Statistics.** In accordance with methods described previously<sup>14,15</sup>, the natural logarithm of the response ratio ( $r = \text{response in elevated [CO}_2\text{]}/\text{response in ambient [CO}_2\text{]}$ ) was used as the metric for analyses and is reported as the mean percentage change ( $100 \times (r - 1)$ ) at elevated [CO<sub>2</sub>]. Consistent with these earlier analyses of multiple species grown under FACE conditions, the responses of different species, cultivars and stress treatments and from different years of the FACE experiments were considered to be independent and suited to meta-analytic analysis<sup>14</sup>.

The meta-analysis was designed to estimate the overall effect of elevated [CO<sub>2</sub>] on the concentration of each nutrient in a particular crop and to determine the significance of this effect relative to a null hypothesis of no change. All tests were conducted as two-sided—not specifying which direction the nutrient concentrations were expected to change under elevated [CO<sub>2</sub>]—to make the analysis as general as possible. Meta-analysis was conducted with a linear mixed model. A random intercept was included for each comparison, representing nutrient level variability unrelated to [CO<sub>2</sub>] that was common to both treatment groups. Additional analyses indicated that the effect of [CO<sub>2</sub>] on zinc concentration in rice was modified by cultivar and amount of nitrogen application, suggesting systematic variations across the pooled analysis of rice, and for these samples it was shown that the effect on zinc concentration was still significant when including interactions terms for cultivar and nitrogen. No other significant modifications of the [CO<sub>2</sub>] effect were identified. We tested whether changes in different nutrients for particular crops were statistically different from each other, as has been described<sup>30</sup>. To address the issue of multiple comparisons when testing for differences between cultivars within a crop, we multiplied the *P* value by the number of independent comparisons. This approach follows the so-called Bonferroni correction and is conservative in the sense of biasing the *P* values high, but still shows that individual test results are significant despite their having been selected from multiple tests.

Parameter estimates were obtained by the restricted maximum-likelihood method, a standard approach for analysing repeated measurement data<sup>29</sup> that, in our case, were of nutrient concentrations at time of harvest. Results for all analyses are reported as the best estimate of percentage changes in the concentration of nutrients along with the 95% confidence intervals associated with each estimate. Two-tailed *P* values are also reported.

When combining our data with previously published data, we defined outliers as pairs in which the difference between an observation at ambient [CO<sub>2</sub>] and elevated [CO<sub>2</sub>] was at least three times the standard deviation from the mean differences for that crop and nutrient type when calculated using all observations. Using this criterion, we excluded a total of two pairs of previously published data from all analyses; these included one observation of iron in rice and one observation of zinc in potato.

**Agricultural methods.** Rice (*Oryza sativa*, 18 cultivars), wheat (*Triticum aestivum*, 8 cultivars), maize (*Zea mays*, 2 cultivars), soybeans (*Glycine max*, 7 cultivars), field peas (*Pisum sativum*, 4 cultivars) and sorghum (*Sorghum bicolor*, 1 cultivar) were grown under FACE conditions during daylight hours. The experiments were conducted in Australia, Japan and the United States between 1998 and 2010. Ambient [CO<sub>2</sub>] ranges were between 363 and 386 p.p.m.; elevated [CO<sub>2</sub>] was between 546 and 584 p.p.m. With the exception of soybeans, each experiment involved multiple cultivars of each crop and more than one set of growing conditions. Each experiment for each cultivar and set of treatments was replicated four times, with the exception of one of the rice sites, for which three replicates were performed. These data are summarized in Table 1, and additional details of the soil and growing conditions, FACE methods and experimental designs have been published for rice<sup>31</sup>, wheat<sup>32</sup>, maize<sup>33</sup>, soybeans<sup>34</sup>, field peas<sup>32</sup> and sorghum<sup>35</sup>.

**Minerals method.** Samples were analysed for minerals by heated closed-vessel digestion/dissolution with nitric acid and hydrogen peroxide followed by quantification with an inductively coupled plasma atomic emission spectrometer<sup>36</sup>. Nitrogen content was measured by flash combustion of the sample coupled with thermal conductivity/infrared detection of the combustion gases (N<sub>2</sub>, NO<sub>x</sub> and CO<sub>2</sub>) with a LECO TruSpec CN Analyzer<sup>37</sup>. Protein values are based on measurement of nitrogen and conversion to protein with the equation below, where  $k = 5.36$  (ref. 38):

$$\text{protein (weight \%)} = k \times \text{nitrogen (weight \%)}$$

For phytic acid determination, a modified version of the method of ref. 39 was used. The accuracy of the method was monitored by the inclusion of tissue standards of known and varying levels of phytic acid<sup>40</sup>.

**Dietary calculations.** The United Nations Food and Agriculture Organization (UNFAO) publishes annual Food Balance Sheets, which provide country-specific data on the quantities of 95 'standardized' food commodities available for human consumption. Data, expressed in terms of dietary energy (kilocalories per person per day) were downloaded for 210 countries and territories with available information for the period 2003–2007 (available at <http://faostat.fao.org>). The percentage of dietary energy available from C<sub>3</sub> grasses (wheat, barley, rye, oats, rice and 'cereals, other' (excluding *Eragrostis tef*)) was calculated globally with estimates weighted by national population size (188 countries available; UN 2011; 2012 revision available at <http://esa.un.org/wpp/>).

Dietary intake data from the UNFAO Food Balance Sheets (to year 2000) and food composition data from the United States Department of Agriculture National Nutrient Database for Standard Reference were used to calculate per-person nutrient intake for 95 food items; these were shared with us with permission<sup>41</sup>. This data set was used to calculate the contribution of each food item to total dietary zinc and iron intake, and the proportions of all food items derived from C<sub>3</sub> grains and legumes were summed to identify countries that are highly dependent on plant sources of iron and zinc (Extended Data Table 5).

30. Schenker, N. & Gentleman, J. F. On judging the significance of differences by examining the overlap between confidence intervals. *Am. Stat.* **55**, 182–186 (2001).
31. Hasegawa, T. A. *et al.* Rice cultivar responses to elevated CO<sub>2</sub> at two free-air CO<sub>2</sub> enrichment (FACE) sites in Japan. *Funct. Plant Biol.* **40**, 148–159 (2013).
32. Mollah, M., Norton, R. & Huzzey, J. Australian Grains Free Air Carbon dioxide Enrichment (AGFACE) facility: design and performance. *Crop Pasture Sci.* **60**, 697–707 (2009).
33. Markelz, R., Strellner, R. & Leahey, A. Impairment of C<sub>4</sub> photosynthesis by drought is exacerbated by limiting nitrogen and ameliorated by elevated CO<sub>2</sub> in maize. *J. Exp. Bot.* **62**, 3235–3246 (2011).
34. Gillespie, K. *et al.* Greater antioxidant and respiratory metabolism in field-grown soybean exposed to elevated O<sub>3</sub> under both ambient and elevated CO<sub>2</sub>. *Plant Cell Environ.* **35**, 169–184 (2012).
35. Ottman, M. J. *et al.* Elevated CO<sub>2</sub> increases sorghum biomass under drought conditions. *New Phytol.* **150**, 261–273 (2001).
36. Sah, R. N. & Miller, R. O. Spontaneous reaction for acid dissolution of biological tissues in closed vessels. *Anal. Chem.* **64**, 230–233 (1992).
37. AOAC Official Method 972.43. in *Official Methods of Analysis of AOAC International, 18th edition, Revision 1, 2006* Ch. 12 5–6 (AOAC International, 2006).
38. Mosse, J. Nitrogen to protein conversion factor for ten cereals and six legumes or oilseeds. A reappraisal of its definition and determination. Variation according to species and to seed protein content. *J. Agric. Food Chem.* **38**, 18–24 (1990).
39. Haug, W. & Lantzsch, H. J. Sensitive method for the rapid determination of phytate in cereals and cereal products. *J. Sci. Food Agric.* **34**, 1423–1426 (1983).
40. Raboy, V. *et al.* Origin and seed phenotype of maize low phytic acid 1-1 and low phytic acid 2-1. *Plant Physiol.* **124**, 355–368 (2000).
41. Wuehler, S. E., Pearson, J. M. & Brown, K. H. Use of national food balance data to estimate the adequacy of zinc in national food supplies: methodology and regional estimates. *Public Health Nutr.* **8**, 812–819 (2005).



**Extended Data Table 1 | Percentage change in nutrient content at elevated [CO<sub>2</sub>] relative to ambient [CO<sub>2</sub>]**

N* (number of pairs)		Zn (µg/g)			Fe (µg/g)			Protein (mg/g)			Phytate (g/100g)		
		%	95% CI	P-value	%	95% CI	P-value	%	95% CI	P-value	%	95% CI	P-value
C3 grasses													
Wheat	64	-9.3	(-12.7,-5.9)	<.0001	-5.1	(-6.5,-3.7)	<.0001	-6.3	(-7.5,-5.2)	<.0001	-4.2	(-7.5,-0.8)	0.009
Rice	31	-3.3	(-5.0,-1.7)	<.0001	-5.2	(-7.6,-2.9)	<.0001	-7.8	(-8.9,-6.8)	<.0001	1.2	(-4.6,7.4)	0.697
C3 legumes													
Field peas	10	-6.8	(-9.8,-3.8)	0.002	-4.1	(-6.7,-1.4)	0.003	-2.1	(-4.0,-0.1)	0.039	-5.8	(-11.5,0.1)	0.055
Soybeans	25	-5.1	(-6.4,-3.9)	<.0001	-4.1	(-5.8,-2.5)	<.0001	0.5	(-0.4,1.3)	0.267	-1.3	(-3.7,1.2)	0.303
C4 grasses													
Maize	4	-5.2	(-10.7,0.6)	0.077	-5.8	(-10.9,-0.3)	0.038	-4.6	(-13.0,4.5)	0.312	-6.1	(-15.0,3.7)	0.215
Sorghum	4	-1.3	(-6.2,3.8)	0.603	1.6	(-5.8,9.7)	0.674	0.0	(-4.9,5.2)	0.993	12.8	(-15.8,51.1)	0.418

\* 'Number of pairs' refers to the number of comparisons in which replicates of a particular cultivar grown at a specific site under one set of growing conditions in one year at elevated [CO<sub>2</sub>] have been pooled and mean nutrient values for these replicates were compared with mean values for identical cultivars under identical growing conditions except grown at ambient [CO<sub>2</sub>]. In most instances, data from four replicates were pooled for each value, meaning that eight experiments were combined for each comparison (see Table 1 for details of experiments).

Extended Data Table 2 | Original data combined with previously published FACE data from studies 3, 4, 6 and 7

N* (number of pairs)		Zn (µg/g)			Fe (µg/g)			Protein (mg/g)		
		%	95% CI	P-value	%	95% CI	P-value	%	95% CI	P-value
<b>C3 grasses</b>										
Wheat	70	-8.8	(-11.9,-5.6)	<.0001	-5.5	(-6.8,-4.1)	<.0001	-6.5	(-7.5,-5.4)	<.0001
Rice	32	-3.1	(-4.8,-1.5)	<.0001	-4.9	(-7.3,-2.6)	<.0001	-8	(-9.0,-6.9)	<.0001
Barley	4	-11.4	(-19.3,-2.7)	0.012	-10.5	(-12.2,-8.7)	<.0001	-11.9	(-13.1,-10.7)	<.0001
<b>C3 legumes</b>										
Field peas	10	-6.8	(-9.8,-3.8)	0.002	-4.1	(-6.7,-1.4)	0.003	-2.1	(-4.0,-0.1)	0.039
Soybeans	25	-5.1	(-6.4,-3.9)	<.0001	-4.1	(-5.8,-2.5)	<.0001	0.5	(-0.4,1.3)	0.267
<b>C3 tubers</b>										
Potato	2	-3.9	(-12.9,6.2)	0.440	2.3	(-3.8,8.7)	0.472	-4.6	(-7.7,-1.4)	<.0001
<b>C4 grasses</b>										
Maize	4	-5.2	(-10.7,0.6)	0.077	-5.8	(-10.9,-0.3)	0.038	-4.6	(-13.0,4.5)	0.312
Sorghum	4	-1.3	(-6.2,3.8)	0.603	1.6	(-5.8,9.7)	0.674	0.0	(-4.9,5.2)	0.993

See Extended Data Table 6 for a list of experiments. Percentage change in nutrient content at elevated [CO<sub>2</sub>] relative to ambient [CO<sub>2</sub>].

\* 'Number of pairs' refers to the number of comparisons in which replicates of a particular cultivar grown at a specific site under one set of growing conditions in one year at elevated [CO<sub>2</sub>] have been pooled and mean nutrient values for these replicates were compared with mean values for identical cultivars under identical growing conditions except grown at ambient [CO<sub>2</sub>]. In most instances, data from four replicates were pooled for each value, meaning that eight experiments were combined for each comparison (see Table 1 for details of experiments).

Extended Data Table 3 | Original data combined with previously published FACE and chamber data from studies 1–10

N* (number of pairs)		Zn (µg/g)			Fe (µg/g)			Protein (mg/g)		
		%	95% CI	P-value	%	95% CI	P-value	%	95% CI	P-value
C3 grasses										
Wheat	78	-9.1	(-12.1,-6.1)	<.0001	-5.9	(-7.8,-4.0)	<.0001	-7.2	(-8.6,-5.8)	<.0001
Rice	32	-3.1	(-4.8,-1.5)	<.0001	-4.9	(-7.3,-2.6)	<.0001	-8	(-9.0,-6.9)	<.0001
Barley	6	-13.6	(-19.3,-7.6)	<.0001	-10.0	(-12.4,-7.4)	<.0001	-15.0	(-19.1,-10.7)	<.0001
C3 legumes										
Field peas	10	-6.8	(-9.8,-3.8)	<.0001	-4.1	(-6.7,-1.4)	0.003	-2.1	(-4.0,-0.1)	0.039
Soybeans	28	-5.0	(-6.1,-3.9)	<.0001	-5.2	(-7.9,-2.5)	<.0001	0.1	(-0.8,0.9)	0.865
C3 tubers										
Potato	5	-10.0	(-20.9,2.4)	0.110	-4.1	(-16.6,10.3)	0.555	-9.7	(-15.9,-3.1)	0.005
C4 grasses										
Maize	4	-5.2	(-10.7,0.6)	0.077	-5.8	(-10.9,-0.3)	0.038	-4.6	(-13.0,4.5)	0.312
Sorghum	7	-0.6	(-4.5,3.4)	0.764	33.8	(-10.2,99.3)	0.153	-5.6	(-12.7,2.1)	0.150

See Extended Data Table 6 for a list of experiments. Percentage change in nutrient content at elevated [CO<sub>2</sub>] relative to ambient [CO<sub>2</sub>].

\* 'Number of pairs' refers to the number of comparisons in which replicates of a particular cultivar grown at a specific site under one set of growing conditions in one year at elevated [CO<sub>2</sub>] have been pooled and mean nutrient values for these replicates were compared with mean values for identical cultivars under identical growing conditions except grown at ambient [CO<sub>2</sub>]. In most instances, data from four replicates were pooled for each value, meaning that eight experiments were combined for each comparison (see Table 1 for details of experiments).



**Extended Data Table 4 | Percentage change in nutrient content at elevated [CO<sub>2</sub>] compared with ambient [CO<sub>2</sub>] for all nutrients**

	C3 grasses						C3 legumes						C4 grasses					
	Wheat			Rice			Field Peas			Soybean			Maize			Sorghum		
	%	95% CI	P-value	%	95% CI	P-value	%	95% CI	P-value	%	95% CI	P-value	%	95% CI	P-value	%	95% CI	P-value
Zinc (ppm)	-9.3	(-12.7,-5.9)	<.0001	-3.3	(-5.0,-1.7)	<.0001	-6.8	(-9.8,-3.8)	<.0001	-5.1	(-6.4,-3.9)	<.0001	-5.2	(-10.7,0.6)	0.077	-1.3	(-6.2,3.8)	0.603
Iron (ppm)	-5.1	(-6.5,-3.7)	<.0001	-5.2	(-7.6,-2.9)	<.0001	-4.1	(-6.7,-1.4)	<.0001	-4.1	(-5.8,-2.5)	<.0001	-5.8	(-10.9,-0.3)	0.038	1.6	(-5.8,9.7)	0.674
Phytate (mg/g)	-4.2	(-7.5,-0.8)	0.009	1.2	(-4.6,7.4)	0.7	-5.8	(-11.5,0.1)	0.055	-1.3	(-3.7,1.2)	0.303	-6.1	(-15.0,3.7)	0.215	12.8	(-15.8,51.1)	0.418
Protein	-6.3	(-7.5,-5.2)	<.0001	-7.8	(-8.9,-6.8)	<.0001	-2.1	(-4.0,-0.1)	0.039	0.5	(-0.4,1.3)	0.267	-4.6	(-13.0,4.5)	0.312	0.0	(-4.9,5.2)	0.993
Mn (ppm)				-7.5	(-12.0,-2.8)	<.0001	-2.5	(-4.2,-0.8)	0.005	-1.4	(-3.5,0.8)	0.204	-4.2	(-10.5,2.5)	0.215	1.7	(-4.5,8.3)	0.596
Mg (%)				-0.9	(-2.3,0.6)	0.24	0.0	(-1.3,1.4)	0.960	-3.5	(-4.3,-2.8)	<.0001	-5.7	(-9.9,-1.3)	0.011	-0.2	(-5.1,4.9)	0.944
Cu (ppm)				-10.6	(-13.8,-7.1)	<.0001	-2.7	(-5.1,-0.3)	0.025	-5.7	(-8.0,-3.4)	<.0001	-9.9	(-19.3,0.7)	0.066	-2.9	(-7.1,1.5)	0.190
Ca (%)				2	(-0.8,4.9)	0.16	-0.5	(-4.2,3.3)	0.787	-5.8	(-7.3,-4.2)	<.0001	-2.7	(-16.9,13.9)	0.734	11.2	(-5.2,30.3)	0.190
S (ppm)				-7.8	(-8.8,-6.8)	<.0001	-2.2	(-3.6,-0.7)	0.003	-2.9	(-3.5,-2.2)	<.0001	2.1	(-2.2,6.7)	0.342	-0.2	(-5.4,5.2)	0.936
K (%)				1.1	(-0.3,2.5)	0.13	2.2	(0.6,3.8)	0.008	0.1	(-0.8,1.0)	0.857	-2.7	(-3.1,-2.2)	<.0001	3.0	(-2.7,9.1)	0.308
B (ppm)				5.1	(1.9,8.4)	0.002	-1.9	(-3.9,0.1)	0.057	-6.4	(-9.1,-3.6)	<.0001	4.9	(-1.0,11.1)	0.107	-0.3	(-9.3,9.6)	0.952
P (%)				-1.0	(-2.4,0.4)	0.160	-3.7	(-6.8,-0.5)	0.023	-0.7	(-2.2,0.9)	0.379	-7.1	(-9.0,-5.1)	<.0001	0.3	(-4.0,4.9)	0.881

Sample sizes for each crop type are identical to those listed in Table 1.

Extended Data Table 5 | Countries whose populations receive at least 60% of dietary iron and/or zinc from C<sub>3</sub> grains and legumes

Country	% Iron from C <sub>3</sub> grains & legumes	% Zinc from C <sub>3</sub> grains & legumes	Population (in thousands)
Afghanistan	78%	78%	31,412
Algeria	76%	79%	35,468
Iraq	74%	83%	31,672
Bangladesh	72%	88%	148,692
Iran, Islamic Rep of	72%	77%	73,974
Pakistan	70%	72%	173,593
Tunisia	70%	77%	10,481
Jordan	69%	73%	6,187
Morocco	69%	78%	31,951
Syrian Arab Republic	67%	71%	20,411
Libya	67%	71%	6,355
Yemen	66%	75%	24,053
Myanmar	65%	81%	47,963
Tajikistan	62%	56%	6,879
India	59%	71%	1,224,614
Egypt	54%	65%	81,121
Indonesia	52%	65%	239,871
Sierra Leone	51%	70%	5,868
Cambodia	49%	68%	14,138
Sri Lanka	46%	69%	20,860
Laos	44%	66%	6,201
Viet Nam	43%	61%	87,848
Total			2,329,612

Source: United Nations Food and Agriculture Organization food balance sheets and 2010 United Nations estimated population.

Extended Data Table 6 | Literature reporting nutrient changes in the edible portion of crops grown at elevated and ambient [CO<sub>2</sub>]

Study	Experimental Method	Associated Citations
1	Growth Chambers	Conroy, J., Seneweera, S. P., Basra, A., Rogers, G. & Nissen-Wooler, B. Influence of rising atmospheric CO <sub>2</sub> concentrations and temperature on growth, yield and grain quality of cereal crops. <i>Australian Journal of Plant Physiology</i> 21, 741-758 (1994).
		Seneweera, S., Milham, P. & Conroy, J. Influence of elevated CO <sub>2</sub> and phosphorus nutrition on the growth and yield of a short-duration rice. <i>Australian Journal of Plant Physiology</i> 21, 281-292 (1994).
		Seneweera, S. P. & Conroy, J. P. Growth, grain yield and quality of rice ( <i>Oryza sativa</i> L.) in response to elevated CO <sub>2</sub> and phosphorus nutrition (Reprinted from <i>Plant nutrition for sustainable food production and environment</i> , 1997). <i>Soil Sci. Plant Nutr.</i> 43, 1131-1136 (1997).
2	Temperature Gradient Tunnels	De la Puente, L. S., Perez, P. P., Martinez-Carrasco, R., Morcuende, R. M. & Del Molino, I. M. M. Action of elevated CO <sub>2</sub> and high temperatures on the mineral chemical composition of two varieties of wheat. <i>Agrochimica</i> 44, 221-230 (2000).
3	Open Top Chambers & FACE	De Temmerman L et al. Effect of climatic conditions on tuber yield ( <i>Solanum tuberosum</i> L.) in the European 'CHIP' experiments. <i>European Journal of Agronomy</i> 17, 243-255 (2002).
		De Temmerman, L., Hacıur, A. & Guns, M. Changing climate and potential impacts on potato yields and quality 'CHIP': introduction, aims and methodology. <i>European Journal of Agronomy</i> 17, 233-242 (2002).
		Fangmeier, A., De Temmerman, L., Black, C., Persson, K. & Vorne, V. Effects of elevated CO <sub>2</sub> and/or ozone on nutrient concentrations and nutrient uptake of potatoes. <i>European Journal of Agronomy</i> 17, 353-368 (2002).
		Högy, P. & Fangmeier, A. Atmospheric CO <sub>2</sub> enrichment affects potatoes: 2. Tuber quality traits. <i>European Journal of Agronomy</i> 30, 85-94 (2009).
4	FACE	Erbs, M. et al. Effects of free-air CO <sub>2</sub> enrichment and nitrogen supply on grain quality parameters and elemental composition of wheat and barley grown in a crop rotation. <i>Agriculture, Ecosystems and Environment</i> 136, 59-68 (2010).
5	Open Top Chambers	Fangmeier, A. et al. Effects of elevated CO <sub>2</sub> , nitrogen supply and tropospheric ozone on spring wheat. I. Growth and yield. <i>Environmental Pollution</i> 91, 381-390 (1996).
		Fangmeier, A., Grütters, U., Högy, P., Vermehren, B. & Jäger, H.-J. Effects of elevated CO <sub>2</sub> , nitrogen supply and tropospheric ozone on spring wheat – II. Nutrients (N, P, K, S, Ca, Mg, Fe, Mn, Zn). <i>Environmental Pollution</i> 96, 43-59 (1997).
		Fangmeier, A. et al. Effects on nutrients and on grain quality in spring wheat crops grown under elevated CO <sub>2</sub> concentrations and stress conditions in the European, multiple-site experiment 'ESPACE-wheat'. <i>European Journal of Agronomy</i> 10, 215-229 (1999).
		Jäger, H.-J., Hertstein, U. & Fangmeier, A. The European Stress Physiology and Climate Experiment – project 1: wheat (ESPACE-wheat): introduction, aims and methodology. <i>European Journal of Agronomy</i> 10, 155-162 (1999).
6	FACE	Högy, P. & Fangmeier, A. Effects of elevated atmospheric CO <sub>2</sub> on grain quality of wheat. <i>Journal of Cereal Science</i> 48, 580-591 (2008).
		Högy, P. et al. Does elevated atmospheric CO <sub>2</sub> allow for sufficient wheat grain quality in the future? . <i>Journal of Applied Botany and Food Quality</i> 82, 114-121 (2009).
		Högy, P. et al. Effects of elevated CO <sub>2</sub> on grain yield and quality of wheat: results from a 3-year free-air CO <sub>2</sub> enrichment experiment. <i>Plant Biology</i> 11, 60-69 (2009).
		Högy, P., Zörb, C., Langenkämper, G., Betsche, T. & Fangmeier, A. Atmospheric CO <sub>2</sub> enrichment changes the wheat grain proteome. <i>Journal of Cereal Science</i> 50, 248-254 (2009).
7	FACE	Kim, H., Lieferrig, M., Miura, S., Kobayashi, K. & Okada, M. Growth and nitrogen uptake of CO <sub>2</sub> -enriched rice under field conditions. <i>New Phytologist</i> 150, 223-229 (2001).
		Kim, H. et al. Effects of free-air CO <sub>2</sub> enrichment and nitrogen supply on the yield of temperate paddy rice crops. <i>Field Crops Research</i> 83, 261-270 (2003).
		Lieferrig, M., Kim, H.-Y., Kobayashi, K. & Okada, M. The impact of elevated CO <sub>2</sub> on the elemental concentrations of field-grown rice grains. <i>Field Crops Research</i> 88, 279-286 (2004).
8	Open Top Chambers	Pleijel, H. et al. Effects of elevated carbon dioxide, ozone and water availability on spring wheat growth and yield. <i>Physiologia Plantarum</i> 108, 61-70 (2000).
		Pleijel, H. & Danielsson, H. Yield dilution of grain Zn in wheat grown in open-top chamber experiments with elevated CO <sub>2</sub> and O <sub>3</sub> exposure. <i>Journal of Cereal Science</i> 50, 278-282 (2009).
9	Open Top Chambers	Prior, S. A., Runion, G. B., Rogers, H. H., Torbert, H. A. Effects of atmospheric CO <sub>2</sub> enrichment on crop nutrient dynamics under no-till conditions. <i>Journal of Plant Nutrition</i> 31, 758-773 (2008).
10	Open Top Chambers	Weigel, H., Manderscheid, R., Jäger, H.-J. & Mejer, G. Effects of season-long CO <sub>2</sub> enrichment on cereals. I. Growth performance and yield. <i>Agriculture, Ecosystems and Environment</i> 48, 231-240 (1994).
		Manderscheid, R., Bender, J., Jäger, H., J. & Weigel, H., J. Effects of season long CO <sub>2</sub> enrichment on cereals. II. Nutrient concentrations and grain quality. <i>Agriculture, Ecosystems &amp; Environment</i> 54, 175-185 (1995).
11	FACE	Yang, L., Wang, Y., Dong, G., Gu, H., Huang, J., Zhu, J., Yang, H., Liu, G., Han, Y. The impact of free-air CO <sub>2</sub> enrichment (FACE) and nitrogen supply on grain quality of rice. <i>Field Crops Research</i> 102, 128-140 (2007).
		Meta-Analyses
		Loladze, I. Rising atmospheric CO <sub>2</sub> and human nutrition: toward globally imbalanced plant stoichiometry? <i>Trends in Ecology and Evolution</i> 17 (10), 457-461 (2002). [Uses data from studies 1, 2, 5, and 10 as well as numerous other studies on non-edible tissues and plants other than food crops].
		McGrath, J. M. & Lobell, D. B. Reduction of transpiration and altered nutrient allocation contribute to nutrient decline of crops grown in elevated CO <sub>2</sub> concentrations. <i>Plant, Cell, &amp; Environment</i> 36, 697-705 (2013). [Uses data from studies 1, 5, and 10 as well as numerous other studies on non-edible tissues and plants other than food crops].
		Duval, B.D., Blankinship, J. C., Dijkstra, P., Hungate, B. A. CO <sub>2</sub> effects on plant nutrient concentration depend on plant functional group and available nitrogen: a meta-analysis. <i>Plant Ecology</i> 213, 505-521 (2012). [Uses data from studies 1,2, 3, 5, 6, and 9 as well as numerous other studies on non-edible tissues and plants other than food crops].



# Coordination of entorhinal–hippocampal ensemble activity during associative learning

Kei M. Igarashi<sup>1</sup>, Li Lu<sup>1</sup>, Laura L. Colgin<sup>2</sup>, May-Britt Moser<sup>1</sup> & Edvard I. Moser<sup>1</sup>

**Accumulating evidence points to cortical oscillations as a mechanism for mediating interactions among functionally specialized neurons in distributed brain circuits<sup>1–6</sup>. A brain function that may use such interactions is declarative memory—that is, memory that can be consciously recalled, such as episodes and facts. Declarative memory is enabled by circuits in the entorhinal cortex that interface the hippocampus with the neocortex<sup>7,8</sup>. During encoding and retrieval of declarative memories, entorhinal and hippocampal circuits are thought to interact via theta and gamma oscillations<sup>4,6,8</sup>, which in awake rodents predominate frequency spectra in both regions<sup>9–12</sup>. In favour of this idea, theta–gamma coupling has been observed between entorhinal cortex and hippocampus under steady-state conditions in well-trained rats<sup>12</sup>; however, the relationship between inter-regional coupling and memory formation remains poorly understood. Here we show, by multisite recording at successive stages of associative learning, that the coherence of firing patterns in directly connected entorhinal–hippocampus circuits evolves as rats learn to use an odour cue to guide navigational behaviour, and that such coherence is invariably linked to the development of ensemble representations for unique trial outcomes in each area. Entorhinal–hippocampal coupling was observed specifically in the 20–40-hertz frequency band and specifically between the distal part of hippocampal area CA1 and the lateral part of entorhinal cortex, the subfields that receive the predominant olfactory input to the hippocampal region<sup>13</sup>. Collectively, the results identify 20–40-hertz oscillations as a mechanism for synchronizing evolving representations in dispersed neural circuits during encoding and retrieval of olfactory–spatial associative memory.**

Neural activity was recorded from entorhinal cortex (EC) and CA1 of 17 rats trained to solve a simplified version of an odour–place association task thought to depend on interfacing of the hippocampus with inputs from the olfactory bulb and the piriform cortex via the EC<sup>13–15</sup> (Fig. 1 and Supplementary Video 1). On each trial, the animal sampled odours in a cue port for 1 s and then, depending on odour identity, ran to either of two cups for food reward (Fig. 1b). Following 3 weeks of training, the percentage of correct trials increased to asymptotic levels, reaching a criterion of two consecutive sessions of 85% correct performance after  $16.8 \pm 0.8$  days (mean  $\pm$  s.e.m.; Fig. 1c; repeated measures analysis of variance (ANOVA):  $F(20, 320) = 95.6$ ,  $P < 0.001$ ). Training-day numbers 2, 6, 10, 14 and 18 were defined as T1–T5, respectively. T5 was post-criterion for all animals.

We first examined spectral activity in CA1 of five well-trained animals. After these animals had reached the 85% performance criterion, electrodes were implanted across the transverse axis of CA1 (Fig. 1a and Extended Data Fig. 1a). Analyses focused on activity during the cue-sampling period, when recall of odour–place associations was expected to be initiated (Fig. 1d and Extended Data Fig. 2). Cue sampling was associated with strong oscillatory activity in the local field potential (LFP) of CA1 (Fig. 1e). The activity change was strongest in the 20–40-Hz frequency range, below the gamma-band frequencies that dominate running behaviour<sup>10,12,16</sup> and in agreement with previous observations in olfaction-based discrimination tasks<sup>15</sup>. Some increase was observed at

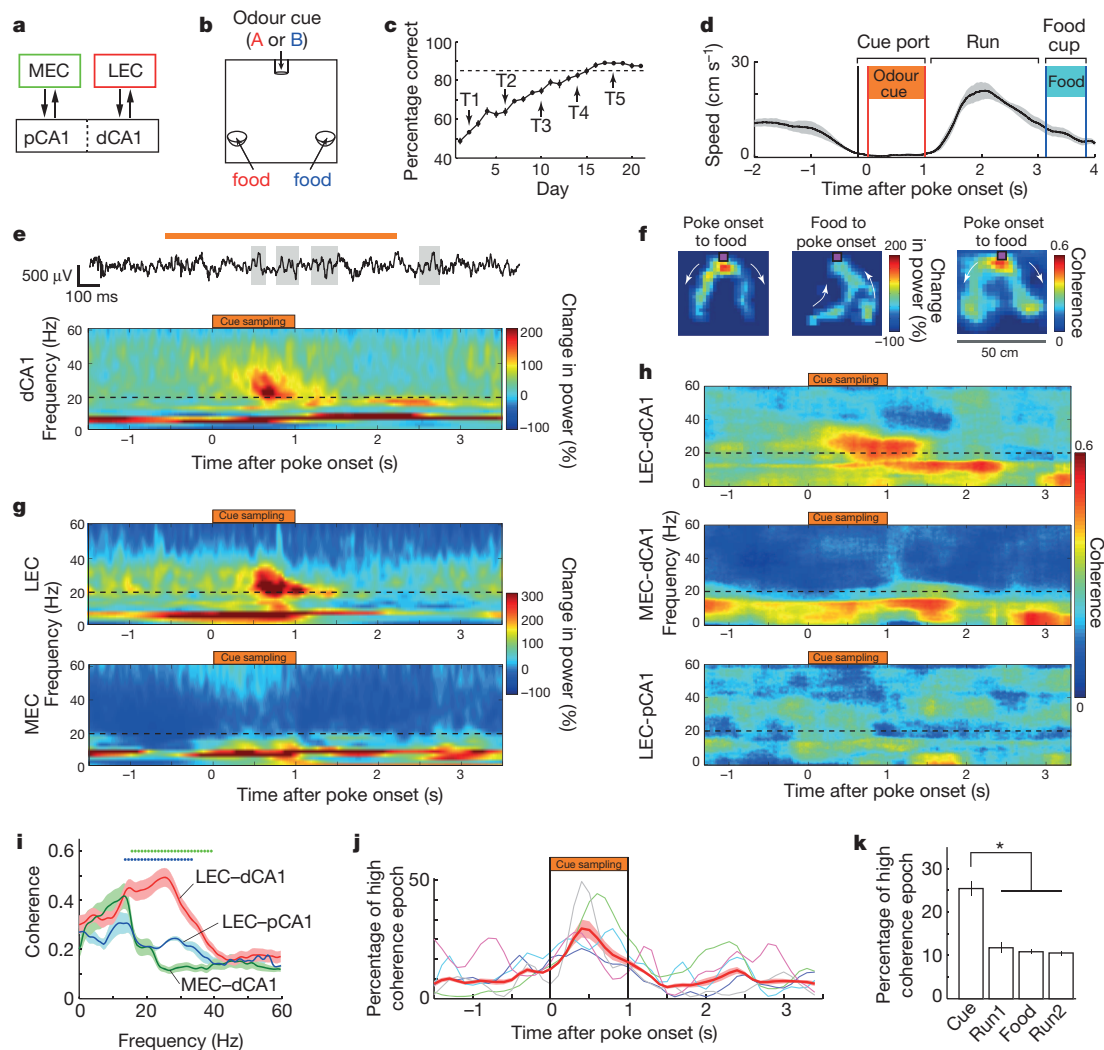
~15–20 Hz, but a 20-Hz cut-off was used to avoid confounds with theta harmonics (~14–18 Hz; Fig. 1e and Extended Data Fig. 3). The power of the 20–40-Hz oscillation increased gradually during the cue sampling interval, peaked during the second half, and dropped back as soon as the animal left the port (Fig. 1e, f, and Extended Data Figs 3 and 4). The magnitude of this oscillation increased from proximal to distal CA1 (dCA1; Extended Data Fig. 5), consistent with the idea that beta-frequency olfactory inputs mediated by the lateral part of the EC (LEC) preferentially reach dCA1<sup>13,15</sup>.

We next asked whether 20–40-Hz oscillations in dCA1 are accompanied by similar LFP activity in EC. In 9 rats, we recorded simultaneously from dCA1 and layer III of either lateral or medial EC (LEC:  $n = 20$  tetrode pairs in 5 rats; MEC:  $n = 16$  tetrode pairs in 4 rats; Extended Data Fig. 1a–c). In all animals that reached the 85% criterion, LEC exhibited strong 20–40-Hz oscillations during cue sampling (Fig. 1g and Extended Data Fig. 3). These oscillations were highly coherent with 20–40-Hz oscillations in dCA1 (Fig. 1h, i). Both power and coherence reached peak values during the second half of the cue interval (Fig. 1g, h). Coherence was increased only rarely during the inter-cue period (Fig. 1f, j, k). The changes were restricted to LEC. No increase in 20–40-Hz oscillations was detected in MEC (Fig. 1g), and oscillations in dCA1 showed stronger coupling with LEC than MEC at all 0.97-Hz frequency bins between 15.6 and 39.0 Hz but not at any lower or higher frequencies (false discovery rate (FDR) corrected for 62 multiple comparisons at 0–60 Hz,  $q < 0.05$ ; Fig. 1h, i). Instead, and in agreement with previous results<sup>12</sup>, the MEC electrodes showed fast gamma oscillations that coupled moderately with dCA1 during the running segments of the task (peak frequency at 72.1 and 61.0 Hz for power and coherence, respectively; Extended Data Fig. 3).

The observed 20–40-Hz coherence between LEC and CA1 could reflect disynaptic LEC–CA3 and CA3–CA1 coupling<sup>12</sup>. If this were the case, LEC–CA1 coherence would be expressed in both proximal CA1 (pCA1) and dCA1, considering that LEC input terminates indiscriminately along the CA3 transverse axis<sup>13</sup>. However, simultaneous recordings from LEC and pCA1 showed no increase in coherence during the cue interval (3 rats; Fig. 1h). At all frequencies from 13.7 to 33.2 Hz, pCA1–LEC coherence was weaker than dCA1–LEC coherence in the first group of rats (FDR corrected,  $q < 0.05$ ; 12 tetrode pairs; Fig. 1i). The lack of coupling between LEC and pCA1, and between MEC and dCA1, suggests that the coherence is mediated by direct LEC–CA1 connections.

We then asked whether and how oscillatory coupling between LEC and dCA1 evolves as the animals learn to use the odour cues to navigate. Tetrodes were targeted simultaneously to dCA1 and layer III of LEC in 5 rats. Once the electrodes were in place, the rats were trained to associate different odours with different reward locations, and LFP as well as individual cell activity were recorded daily from each region, from T1 to T5. Learning was accompanied by a steady increase in 20–40-Hz coherence between LEC and dCA1 (Fig. 2a–c and Extended Data Fig. 6g; repeated measures ANOVA for all 20 combinations of recording pairs:  $F(4, 76) = 35.7$ ,  $P < 0.001$ ; one recording pair per animal:  $F(4, 16) = 17.9$ ,  $P < 0.001$ ). At T5, LEC–dCA1 coherence was stronger

<sup>1</sup>Kavli Institute for Systems Neuroscience and Centre for Neural Computation, Norwegian University of Science and Technology, Olav Kyrres gate 9, MTFS, 7491 Trondheim, Norway. <sup>2</sup>Center for Learning and Memory, The University of Texas at Austin, Austin, Texas 78712-0805, USA.



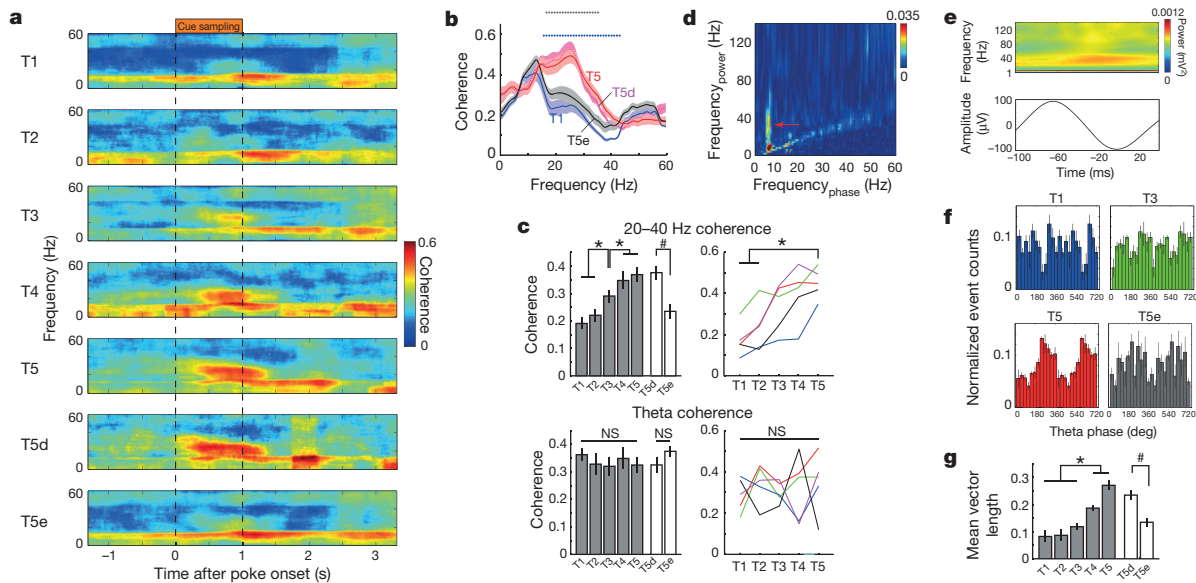
**Figure 1 | 20–40-Hz coupling between LEC and dCA1 during successful odour discrimination.** **a**, Box diagram showing recording locations and their connections. **b**, Task chamber. **c**, Correct choice rate (mean  $\pm$  s.e.m.) as a function of training day (all animals). **d**, Mean speed  $\pm$  s.e.m. (shading) before, during and after odour cue. **e**, Top, example trace of local field potential (LFP) from dCA1. Orange bar, cue period. Grey shadings indicate 20–40-Hz oscillations with power  $>2$ s.d. above the mean. Bottom, time-resolved power spectrum averaged across dCA1 tetrodes (all animals). LFP power was normalized to pre-cue power and aligned to cue onset ( $t = 0$ ). Note increase in 20–40-Hz oscillation during cue sampling ( $t(8) = 3.3$ ,  $P = 0.01$ , paired  $t$ -test). **f**, Map of task chamber showing spatial distribution of 20–40-Hz power change between cue sampling and food retrieval (left) and between food retrieval and cue sampling (middle). Right, spatial distribution of 20–40-Hz LEC–dCA1 coherence (cue sampling to food retrieval). **g**, Time-resolved power spectra averaged across all tetrodes of all rats in LEC ( $t(9) = 4.8$ ,  $P < 0.001$ ) and MEC ( $t(7) = 1.3$ ,  $P = 0.24$ ).

than at T1 for each 0.97-Hz frequency bin between 15.6 and 43.0 Hz (Fig. 2b, c; FDR corrected for 62 multiple comparisons at 0–60 Hz,  $q < 0.05$ ). Coherence in the theta-band showed no change (Fig. 2c;  $F(4, 76) = 0.79$ ,  $P = 0.54$ ). The development of interregional coupling was accompanied by gradual cross-frequency coupling between theta and 20–40-Hz oscillations in dCA1 (Fig. 2d–g; repeated measures ANOVA for mean vector length of 20–40-Hz event distribution across theta phase:  $F(4, 36) = 36.6$ ,  $P < 0.001$ ), supporting earlier observations in CA3 in an item-context association task<sup>17</sup>. In LEC, by contrast, 20–40-Hz and theta oscillations were phase-coupled from the start of training (Extended Data Fig. 7). In both regions, the power of the 20–40-Hz oscillations remained unaltered (Extended Data Fig. 6a–d), suggesting that learning is primarily associated with increased coupling of pre-existing 20–40-Hz rhythms.

**h**, Time-resolved coherence spectra averaged over all recording pairs in LEC–dCA1, MEC–dCA1 and LEC–pCA1. Odour-evoked 20–40-Hz coupling is apparent in LEC–dCA1 (cue versus pre-cue:  $t(19) = 4.4$ ,  $P < 0.001$ ) but not in MEC–dCA1 or LEC–pCA1. **i**, Mean coherence  $\pm$  s.e.m. (shading) as a function of frequency for all LEC–dCA1, MEC–dCA1 and LEC–pCA1 recording pairs in all animals. Green and blue dots at the top indicate frequencies with significant LEC–dCA1 versus MEC–dCA1 and LEC–dCA1 versus LEC–pCA1 differences, respectively ( $q < 0.05$ , FDR corrected, all tetrode pairs). **j**, Percentage of high 20–40-Hz LEC–dCA1 coherence as a function of trial time (mean  $\pm$  s.e.m. of all recording pairs). 100-ms blocks with coherence exceeding 2s.d. of coherence across the entire session were defined as high-coherence epochs. Pale coloured lines, individual rats. **k**, Percentage of high-coherence epochs during cue sampling (cue), running (run1, to food cup; run2, back to cue port), and at food cup (food) (repeated-measures ANOVA:  $F(3, 57) = 36.0$ ,  $P < 0.001$ ; \* $P < 0.001$ , Bonferroni post-hoc test).

To determine whether coupling of 20–40-Hz oscillators in LEC and dCA1 is necessary for successful performance, we assessed activity at T5 on error trials ('T5e'; Supplementary Video 1). Results were compared to an identical number of correct trials from the same session (down-sampled correct trials, 'T5d'). The coherence of LEC and dCA1 oscillations in the 20–40-Hz band decreased significantly on error trials (Fig. 2a–c; paired  $t$ -test for all combinations of recording pairs,  $t(19) = 9.81$ ,  $P < 0.001$ ; for one pair per animal:  $t(4) = 5.0$ ,  $P = 0.007$ ). There was also a decrease in cross-frequency coupling in dCA1 (Fig. 2f, g;  $t(9) = 6.30$ ,  $P < 0.001$ ). These findings suggest that 20–40-Hz coupling of dCA1 and LEC is necessary for successful odour-based navigation.

The emergence of 20–40-Hz coupling was reflected in individual dCA1 and LEC cells. A large proportion of the cue-port cells<sup>18</sup> in dCA1 were



**Figure 2 | Evolution of 20–40-Hz coupling between LEC and dCA1 during odour–place learning.** **a–c**, Coherence between LEC and dCA1 evolves in parallel with learning (average of 20 recording pairs in 5 rats). **a**, Time-resolved LEC–dCA1 coherence spectra. **b**, Mean coherence  $\pm$  s.e.m. as a function of frequency. Top, blue and grey dots indicate frequencies at which T1–T5 and T5d–T5e differences, respectively, were significant ( $q < 0.05$ , FDR corrected). **c**, Coherence for 20–40-Hz (top) and theta frequencies (bottom) from T1 to T5 and on T5 error trials (T5e). T5d, correct T5 trials after down-sampling to same number of trials as T5e. Left, mean coherence  $\pm$  s.e.m. (\* $P < 0.05$ ,

Bonferroni post-hoc test; # $P < 0.05$ , paired  $t$ -test); right, individual animals (colour-coded). **d**, Example cross-frequency coherence plot showing that 20–40-Hz oscillatory power (y-axis) was modulated by theta phase (x-axis) in dCA1 at T5. Coupling strength is colour-coded. **e**, Top, time-resolved power spectrum averaged across all theta cycles with 20–40-Hz oscillations ( $t = 0$ , theta trough). Bottom, averaged unfiltered theta cycle. **f**, Theta phase distribution of 20–40-Hz oscillation maxima for LFP from dCA1 (mean  $\pm$  s.e.m.;  $0^\circ$ , trough). **g**, Mean vector length of distributions in **f** (T1 to T5 and T5e). \* $P$  and # $P$  as in **c**.

place cells (Extended Data Fig. 8). The proportion of cue-port cells with spike activity significantly phase-locked to local 20–40-Hz oscillations increased from 19% to 42% in dCA1, and from 22% to 65% in LEC (T1 and T5, respectively; Extended Data Fig. 9; binomial tests with Bonferroni correction,  $P < 0.005$ ). In dCA1, the phase-locking was reduced on error trials (T5e versus T5d:  $P = 0.03$ , LEC:  $P = 0.097$ ). There was also an increase from T1 to T5 in cross-regional phase locking (from 7.9% to 17.4% for dCA1 cells versus 20–40-Hz LEC oscillations; from 13.3% to 24.4% for LEC cells versus dCA1 oscillations; both  $P < 0.005$ ; Extended Data Fig. 9). The match between oscillatory coupling in LFP and individual neurons suggests that the change occurred locally in LEC and dCA1.

To establish whether oscillatory coupling evolved as a reflection of changes in the animals' motor behaviour, we monitored the path of the rats. There was no change in position and head direction before, during or after cue sampling (Extended Data Fig. 2a–f). Power and coherence in dCA1 and LEC, as well as 20–40-Hz–theta cross-frequency coupling, were significantly larger during the cue interval than during immobility at other times (Extended Data Fig. 4), suggesting that the 20–40-Hz activity at the cue port is independent of relationships between speed and gamma frequency<sup>19</sup>.

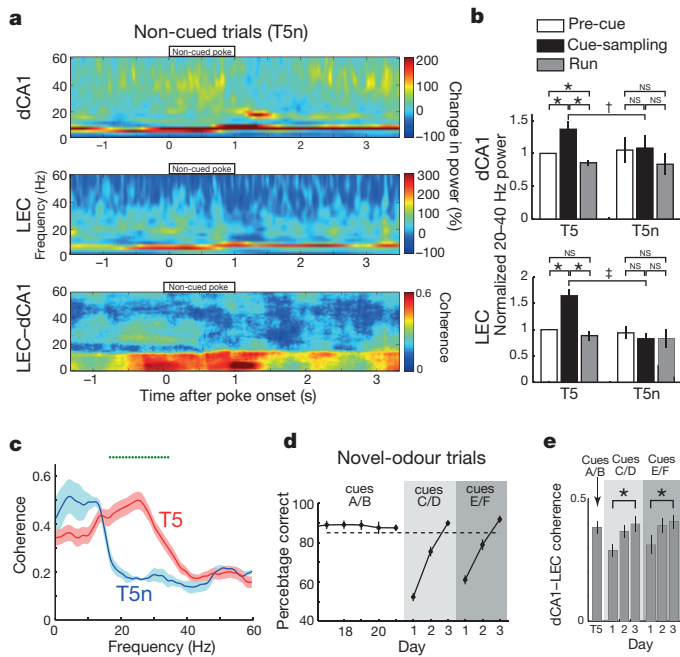
Could the coupling reflect other non-specific behavioural changes, such as enhanced attention to the odour stimuli, which might be accompanied by increased power and coherence of beta-gamma oscillations<sup>20,21</sup>? To dissociate memory from attention, we conducted two control tasks, each designed to maintain attention but disrupt memory-based performance. In both tasks, attention was operationalized by monitoring the sniffing frequency of the rat during the cue interval at T5 (Extended Data Fig. 2g–i). In the first test, we added, for each animal, a set of non-cued trials (T5n) in which only clean air was delivered at the odour port. The increase in 20–40-Hz power was blocked on these trials (Fig. 3a–c; power: dCA1:  $t(8) = 3.50$ ,  $P = 0.008$ ; LEC:  $t(9) = 8.40$ ,  $P < 0.001$ ), as was the increase in coherence (reduced compared to T5 at all frequencies between 16.6 and 34.2 Hz). Sniffing was maintained (Extended Data Fig. 2n–o). In the second test, we trained 4 of the animals with two

additional odour pairs (C/D and E/F). After each switch of odour pairs, performance dropped temporarily, and then reverted back to criterion within three days (Fig. 3d). This development was paralleled by a significant drop in 20–40-Hz coupling between LEC and dCA1 on day 1 (paired  $t$ -test for T5 versus day 1:  $t(15) = 2.19$ ,  $P = 0.04$ ) and complete recovery by day 3 ( $F(2, 30) = 4.05$ ,  $P = 0.03$  for C/D and  $F(2, 30) = 4.54$ ,  $P = 0.02$  for E/F) (Fig. 3e and Extended Data Fig. 6e, f). There was no change in sniffing frequency during the change of odour pairs ( $t(3) < 0.5$ ,  $P > 0.6$ ; Extended Data Fig. 2p). These two series of observations suggest that 20–40-Hz coupling was not primarily a reflection of attention.

In a final set of analyses, we examined if the changes in spike timing were accompanied by changes in the subset of cells recruited by each odour. In both dCA1 and LEC, a substantial fraction of the cells developed selectivity for one of the odours (Fig. 4a–c). 'Selectivity' was expressed for each cell as the difference in firing rate during each odour divided by the sum of these rates. On individual trials, average selectivity increased during the cue period, mirroring the development of 20–40-Hz oscillations (Fig. 4a). The development of unique representations for left and right odours was confirmed by population vector analysis (Fig. 4b, c) as well as receiver-operating characteristic (ROC) analysis (Extended Data Fig. 10). Significant changes in selectivity were not observed in pCA1 or MEC (data not shown). The selectivity of the dCA1 and LEC representations was nearly abolished on error trials and non-cued trials (Fig. 4a–c and Extended Data Fig. 10). When a new odour pair was introduced, the representations became non-selective but reverted to fully discriminative levels by day 3, in parallel with correct behavioural choices (Fig. 4f), suggesting that expression of odour-specific representations in dCA1 and LEC is necessary for successful use of odour cues to guide navigation.

The development of odour-selective representations in dCA1 and LEC was strongly correlated with 20–40-Hz coupling between the two regions (Fig. 4d; T1–T5  $\times$  5 animals, all  $r > 0.48$ ,  $P < 0.05$ ). LEC selectivity, 20–40-Hz coupling and task performance developed in parallel whereas selectivity in dCA1 lagged behind (Fig. 4e). A repeated measures ANOVA on normalized values for each variable, with T1 set to 0



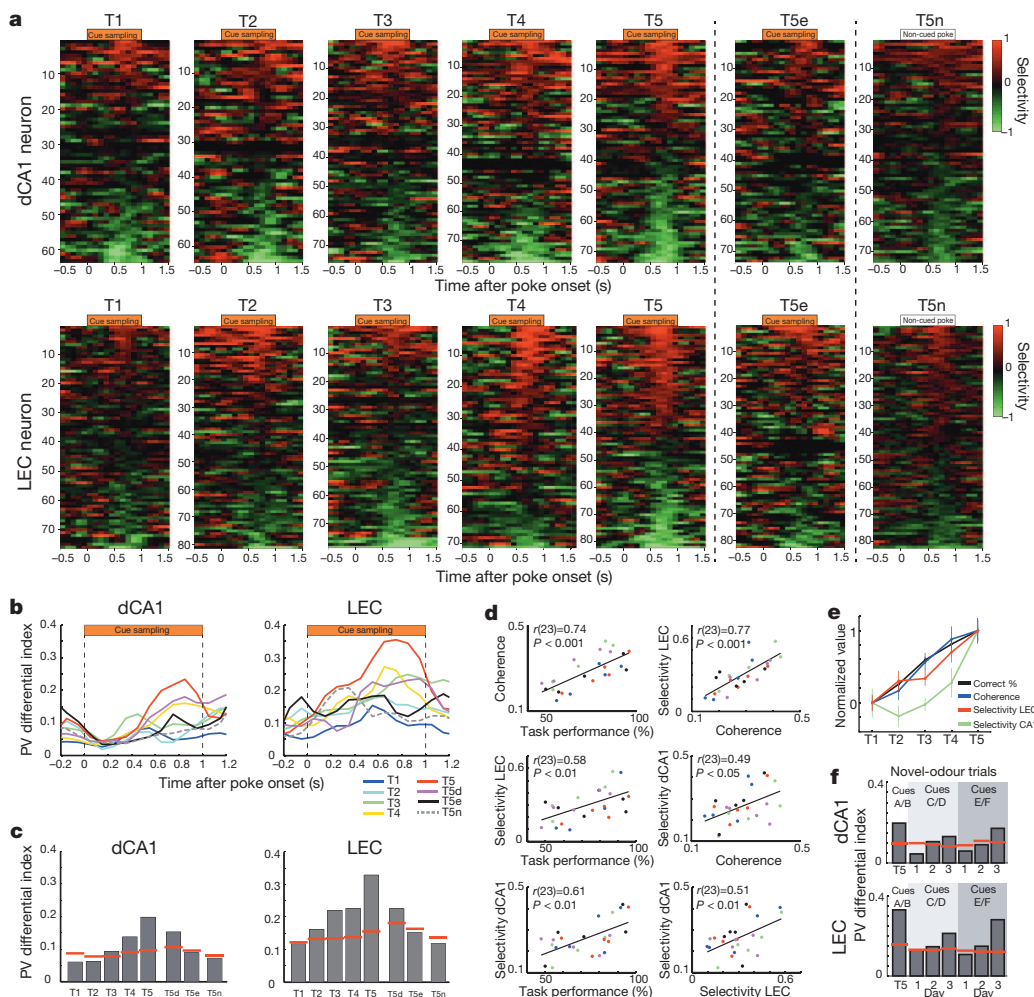


**Figure 3 | 20–40-Hz coupling on non-cued trials and novel-odour trials.**

**a**, Time-resolved power and coherence spectra on non-cued (T5n) trials, averaged across all tetrodes and animals and normalized to the pre-cue period. Top, power for dCA1; middle, power for LEC; bottom, time-resolved coherence spectra for LEC–dCA1. **b**, Normalized power of 20–40-Hz oscillations for cued (T5) and non-cued (T5n) trials ( $*P < 0.05$ ;  $\dagger P < 0.01$ ;  $\ddagger P < 0.001$ ). **c**, Mean of T5 and T5n coherence spectra for the cue period. Dots indicate frequencies with significant difference ( $q < 0.05$ , FDR corrected). **d**, Task performance with novel odour pairs (C/D, followed by E/F). **e**, Coherence averaged across the 20–40-Hz band during cue sampling on successive days with novel odours.

and T5 to 1, showed a significant main effect for trials T2–T4 ( $F(3, 16) = 9.16$ ,  $P = 0.001$ ) and no interaction with trial number:  $F(6, 32) = 0.69$ ,  $P = 0.66$ ). Bonferroni post-hoc tests were non-significant for all combinations of task performance, LEC–dCA1 coupling and LEC selectivity, but significant at  $P < 0.01$  for all comparisons with dCA1 selectivity, suggesting that LEC triggers the changes in spike timing in the hippocampus.

The observation of 20–40-Hz oscillations in the hippocampus is consistent with reports of such activity during odour discrimination<sup>15</sup> and free running in novel<sup>22</sup> or familiar<sup>12</sup> environments. However, the interaction of these rhythms with external cortical networks and the function of such interactions in memory have remained elusive. Using a hippocampus-dependent associative discrimination task, we have shown that learning is accompanied by coupling of 20–40-Hz oscillations in



**Figure 4 | Development of odour-specific representations in dCA1 and LEC cells.**

**a**, Response of all cue port cells in dCA1 at T1–T5, T5e and T5n. Each row shows data for one cell around the time of odour sampling. Top, dCA1 cells; bottom, LEC cells. Selectivity for odour cues is expressed for each cell as the difference in firing rate in the presence of left- and right-associated odours, divided by the sum of these rates. Selectivity is colour-coded (–1 and +1 indicate complete selectivity for right and left, respectively). Cells are sorted according to selectivity. **b**, Population vector (PV) analysis across 100-ms time bins for all dCA1 and LEC cells with cue-port activity at T1–T5, T5e and T5n. T5d as in Fig. 2c. Correlations on left and right trials are scaled and shown as a PV differential index (PDI; 0 indicates identical left–right pattern). **c**, Averaged PDIs for second half of odour interval. Red lines indicate 95th percentiles from shuffled distributions at each time point. **d**, Correlation between task performance (correct choice rate; Fig. 1c), LEC–dCA1 coherence at 20–40 Hz (Fig. 2c), and mean selectivity of dCA1 and LEC cells (Fig. 4a). Each dot represents averaged data for a single animal at one time point (5 rats  $\times$  5 time points). Colours indicate rat identities. **e**, Development of task performance, coherence and selectivity in dCA1 and LEC. Variables are normalized onto a scale from 0 (T1) to 1 (T5) (mean  $\pm$  s.e.m.;  $n = 5$  rats). **f**, PDIs for trials with novel odour pairs.

connected cell populations of LEC and dCA1, and that, in each of these areas, such coupling coincides with the formation of unique odour representations. Error trials and change of odour contingencies were invariably accompanied by reduced coupling and reduced ensemble selectivity, suggesting that the changes are necessary for successful retrieval or retrieval-based navigation. The results identify 20–40-Hz coupling as a potential mechanism for the evolution of functional circuits during encoding of associative memory in entorhinal–hippocampal systems. Coherent firing between LEC and CA1, timed by 20–40-Hz oscillations, may be a prerequisite for the formation of distinct representations for differentially associated odour cues, both because coupling provides sufficient coincidence of pre- and postsynaptic activity for synaptic strengthening to take place<sup>3,23</sup> and because coincident firing among afferent neurons facilitates such strengthening<sup>24</sup>. Interregional coupling may facilitate the pattern-completion processes required for successful reactivation of newly formed entorhinal and hippocampal cell ensembles.

20–40-Hz oscillations occur widely across the cortex<sup>20,25</sup>. During olfactory learning, 20–40-Hz oscillations may coordinate activity across distributed areas spanning from the olfactory bulb to the LEC and the hippocampus. Increases in 20–40-Hz rhythms in the olfactory bulb and olfactory cortex have been observed after olfactory learning<sup>15,26</sup>, in parallel with 20–40-Hz field oscillations in the hippocampus<sup>15</sup>. Widespread 20–40-Hz oscillations have also been observed to emerge at late stages of habit learning in the ventromedial striatum<sup>27</sup>. Because 20–40-Hz oscillations can maintain synchrony with larger conduction delays than faster gamma oscillations<sup>3,28</sup>, they are particularly suitable for linking activity across widely distributed brain regions such as those participating in olfaction-based memory formation. The strong coincidence of entorhinal–hippocampal 20–40-Hz coupling and the emergence of olfactory neural representations in LEC and dCA1, in parallel with behavioural learning, points to 20–40-Hz oscillations as a key ingredient of the mechanism for induction and expression of long-term memory in distributed cortical circuits.

## METHODS SUMMARY

Seventeen male Long-Evans rats were trained to perform an odour–place association task. The rats sampled odours for 1 s in an odour port, after which food was delivered in one of two food cups whenever the rat chose the food cup associated with the presented odour. The food cups were placed at the other end of the test chamber. Behavioural testing was controlled by custom software triggering the valves of the olfactometer as well as food-delivery.

The rats were implanted either with a ‘hyperdrive’ carrying 14 tetrodes targeted to cover almost the entire proximodistal extent of dorsal CA1 ( $n = 5$  rats), or with two ‘microdrives’ carrying 4 tetrodes targeted, respectively, to the distal part of CA1 and either the anterodorsal part of LEC (5 rats) or the dorsal part of MEC (4 rats). In three rats, microdrives were targeted simultaneously to LEC and the proximal part of CA1. A multi-tetrode parallel recording system (Neuralynx or Axona) was used to obtain local field potentials and spike activity from individual cells, to monitor animal behaviour, and to time odour and food delivery. In rats with implants in dCA1 and LEC, implantation and tetrode turning were performed prior to the main part of the behavioural training (T1 to T5).

Analyses were focused on the ‘cue sampling period’ during which the rats stood still and actively sampled odour cues at the cue port. Because all animals were trained and tested similarly, there was no randomization or blinding. The experiments were performed in accordance with the Norwegian Animal Welfare Act and the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 4 July 2013; accepted 17 February 2014.**

**Published online 16 April 2014.**

- Freeman, W. J. Spatial properties of an EEG event in the olfactory bulb and cortex. *Electroencephalogr. Clin. Neurophysiol.* **44**, 586–605 (1978).

- Gray, C. M., Konig, P., Engel, A. K. & Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–337 (1989).
- Singer, W. Synchronization of cortical activity and its putative role in information processing and learning. *Annu. Rev. Physiol.* **55**, 349–374 (1993).
- Buzsáki, G. *Rhythms of the Brain* (Oxford University Press, 2006).
- Fries, P. Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu. Rev. Neurosci.* **32**, 209–224 (2009).
- Colgin, L. L. Mechanisms and functions of theta rhythms. *Annu. Rev. Neurosci.* **36**, 295–312 (2013).
- Squire, L. R. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* **99**, 195–231 (1992).
- Buzsáki, G. & Moser, E. I. Memory, navigation and theta rhythm in the hippocampal–entorhinal system. *Nature Neurosci.* **16**, 130–138 (2013).
- Buzsáki, G., Leung, L. W. & Vanderwolf, C. H. Cellular bases of hippocampal EEG in the behaving rat. *Brain Res.* **287**, 139–171 (1983).
- Bragin, A. et al. Gamma (40–100 Hz) oscillation in the hippocampus of the behaving rat. *J. Neurosci.* **15**, 47–60 (1995).
- Chrobak, J. J. & Buzsáki, G. Gamma oscillations in the entorhinal cortex of the freely behaving rat. *J. Neurosci.* **18**, 388–398 (1998).
- Colgin, L. L. et al. Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature* **462**, 353–357 (2009).
- Witter, M. P. & Amaral, D. G. in *The Rat Nervous System* (ed. Paxinos, G.) 635–704 (Elsevier Academic, 2004).
- Day, M., Langston, R. & Morris, R. G. Glutamate-receptor-mediated encoding and retrieval of paired-associate learning. *Nature* **424**, 205–209 (2003).
- Martin, C., Beshel, J. & Kay, L. M. An olfacto-hippocampal network is dynamically involved in odor-discrimination learning. *J. Neurophysiol.* **98**, 2196–2205 (2007).
- Montgomery, S. M. & Buzsáki, G. Gamma oscillations dynamically couple hippocampal CA3 and CA1 regions during memory task performance. *Proc. Natl Acad. Sci. USA* **104**, 14495–14500 (2007).
- Tort, A. B., Komorowski, R. W., Manns, J. R., Kopell, N. J. & Eichenbaum, H. Theta-gamma coupling increases during the learning of item-context associations. *Proc. Natl Acad. Sci. USA* **106**, 20942–20947 (2009).
- Eichenbaum, H., Kuperstein, M., Fagan, A. & Nagode, J. Cue-sampling and goal-approach correlates of hippocampal unit activity in rats performing an odor-discrimination task. *J. Neurosci.* **7**, 716–732 (1987).
- Ahmed, O. J. & Mehta, M. R. Running speed alters the frequency of hippocampal gamma oscillations. *J. Neurosci.* **32**, 7373–7383 (2012).
- Fries, P., Reynolds, J. H., Rorie, A. E. & Desimone, R. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* **291**, 1560–1563 (2001).
- Buschman, T. J. & Miller, E. K. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortex. *Science* **315**, 1860–1862 (2007).
- Berke, J. D., Hetrick, V., Breck, J. & Greene, R. W. Transient 23–30 Hz oscillations in mouse hippocampus during exploration of novel environments. *Hippocampus* **18**, 519–529 (2008).
- Bi, G. Q. & Poo, M. M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18**, 10464–10472 (1998).
- McNaughton, B. L., Douglas, R. M. & Goddard, G. V. Synaptic enhancement in fascia dentata: cooperativity among coactive afferents. *Brain Res.* **157**, 277–293 (1978).
- Engel, A. K. & Fries, P. Beta-band oscillations — signalling the status quo? *Curr. Opin. Neurobiol.* **20**, 156–165 (2010).
- Ravel, N. et al. Olfactory learning modifies the expression of odour-induced oscillatory responses in the gamma (60–90 Hz) and beta (15–40 Hz) bands in the rat olfactory bulb. *Eur. J. Neurosci.* **17**, 350–358 (2003).
- Howe, M. W., Atallah, H. E., McCool, A., Gibson, D. J. & Graybiel, A. M. Habit learning is associated with major shifts in frequencies of oscillatory activity and synchronized spike firing in striatum. *Proc. Natl Acad. Sci. USA* **108**, 16801–16806 (2011).
- Kopell, N., Ermentrout, G. B., Whittington, M. A. & Traub, R. D. Gamma rhythms and beta rhythms have different synchronization properties. *Proc. Natl Acad. Sci. USA* **97**, 1867–1872 (2000).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank A. M. Amundsgård, K. Haugen, K. Jenssen, E. Kråkvik, R. Skjerpeng and H. Waade for technical assistance, and M. Witter and members of the Moser laboratory for discussions. This work was supported by two Advanced Investigator grants from the European Research Council (‘CIRCUIT’, Grant Agreement no. 232608; ‘ENSEMBLE’, Grant Agreement no. 268598), the Kavli Foundation, the Centre of Excellence scheme of the Research Council of Norway (Centre for the Biology of Memory and Centre for Neural Computation), the Mishima Kaiun Memorial Foundation, and the Japan Society for the Promotion of Science.

**Author Contributions** K.M.I., M.-B.M. and E.I.M. designed experiments and analyses; K.M.I. and L.L. performed the experiments; K.M.I. performed the majority of the analyses, with input from L.L.C., M.-B.M. and E.I.M.; K.M.I. and E.I.M. wrote the paper with input from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.I.M. ([edvard.moser@ntnu.no](mailto:edvard.moser@ntnu.no)) or K.M.I. ([kei.igarashi@ntnu.no](mailto:kei.igarashi@ntnu.no)).



## METHODS

**Subjects.** Data were obtained from 17 male Long Evans rats. In 5 of the rats, the tetrodes were implanted along almost the entire proximodistal extent of CA1 using an oval-bundle 'hyperdrive' consisting of 14 independently movable tetrodes<sup>29</sup>. Five of the rats were implanted with two microdrives carrying a single bundle of 4 tetrodes targeted, respectively, to anterodorsal LEC and distal CA1. Four rats were implanted with two microdrives targeted, respectively, to dorsal MEC and distal CA1. Finally, three rats were implanted with two microdrives aimed, respectively, at anterodorsal LEC and proximal CA1. Sample size was determined as the minimal number of animals that would provide statistical power to detect a group difference. We used all animals in which tetrodes hit the target regions. We did not use randomization to allocate animals to experimental treatments and the experimenters were not blind to the identity of the experimental groups.

All animals were 3–5 months old (450–600 g) at the time of implantation. After surgery, they were housed individually in Plexiglas cages (45 × 44 × 30 cm) in a humidity- and temperature-controlled environment. The rats recovered from surgery for at least one week before odour–place training. During training, the rats were put on a food deprivation regimen that kept them at 80–90% of their free-feeding body weight. The animals were kept on a 12-h light/12-h dark schedule. All testing occurred during the dark phase. The experiments were performed in accordance with the Norwegian Animal Welfare Act and the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes.

**Electrode implantation and surgery.** Tetrodes were constructed from four twisted 17 µm polyimide-coated platinum-iridium (90%–10%) wires (California Fine Wire). The electrode tips were plated with platinum to reduce electrode impedances to between 150 and 300 kΩ at 1 kHz. During implantation surgery, the animals were anaesthetized with isoflurane. Air flow was kept at 1.2 l min<sup>-1</sup> with 0.5–3% (v/v) isoflurane as determined by physiological monitoring. Local anaesthetic (Xylocaine) was applied on the skin before making the incision. Holes were drilled on the dorsal skull, after which the rat was implanted with either two microdrives or a hyperdrive. Hyperdrives for CA1 were implanted in the right hemisphere. Microdrives for dCA1/pCA1 and LEC/MEC were implanted in the right and left hemispheres, respectively. Bihemispheric implants were used because layer III EC cells have both ipsi- and contralateral projections to CA1<sup>13,30</sup>. Hyperdrive tetrodes were implanted vertically between AP 3.4 and 4.9 mm, and ML 2.1 and 3.7 mm, relative to bregma in order to cover as much of the proximodistal axis of CA1 as possible<sup>29</sup>. For microdrive implants, one drive was implanted either in the distal or the proximal part of CA1 (AP 4.2, ML 3.3 and AP 3.2, ML 4.0, respectively, distance from brain surface 1.5 mm, tetrode tips pointing medially at an angle of 30° in the coronal plane). The other drive was implanted in either dorsal anterior LEC (AP 6.1, ML 5.0, distance from brain surface 4.5 mm, with tetrode tips pointing laterally at an angle of 16° in the coronal plane) or in the dorsomedial quarter of the MEC (4.5 mm lateral to midline, 0.2–0.3 mm anterior to the transverse sinus, distance from brain surface 1.5 mm, tetrode tips pointing in the anterior direction at an angle of 10° in the sagittal plane). Hyperdrives and microdrives were fixed to the skull with jewellers' screws and dental cement. Two screws inserted above the cerebellum were used as reference and ground electrodes during recordings. For recording of sniffing behaviour, a temperature sensor (a 0.005-inch, Teflon-coated thermocouple, Omega, 5TC-TT-K-36-36) was implanted in the right nostril through a hole made in the dorsal skull. The implant was secured with dental cement. Thermocouples were inserted into the nasal passage through the nasal bone just rostral to the turbinates.

**Tetrode placement.** Over the course of ~2–3 weeks, the tetrodes were lowered in steps of 200 µm or less until well-separated, large-amplitude, low-frequency neurons appeared in the CA1 pyramidal cell layer or in LEC layer III at a depth of ~7.0 mm. In MEC, tetrodes were targeted at layer III at a depth of ~2.0–4.0 mm. Tetrode positions were maintained across sessions, except in LEC and MEC, where the tetrodes were lowered 100 µm to record from a new set of neurons after the recordings at time point T5 were completed. All tetrode locations were verified histologically at the end (Extended Data Fig. 1). Only animals with tetrodes in the regions of interest were used in this study.

**Behavioural task.** Cue–place association task. All rats were tested in an odour place association task (Supplementary Video 1). The test box (50 × 50 × 40 cm) contained a cylindrical port for odour delivery ('cue port', 2.9 cm internal diameter, i.d.), and two circular food cups for food delivery (4.5 cm i.d.). Entry and exit from odour ports and food cups were detected via an infrared photo-beam located inside the odour port and the food cups. Three pairs of monomolecular odorants (Sigma-Aldrich) were used as odour cues. Two odours (A/B) were used during the initial stage, of the task (A = isoamyl acetate (banana odour) / B = α-pinene (pine tree odour)). For subsequent stages with novel odours, we used *t*-anethol (anis odour) as C and geraniol (rose odour) as D, and undecanoic γ-lactone (peach odour) as E and benzaldehyde (almond odour) as F. All odours were diluted in mineral oil (Sigma) to a final concentration of 5% (v/v).

Odour cues were delivered using an olfactometer (PHM-275, Med Associates, St. Albans, VT, USA) housed in a shielded box and located next to the behavioural box. Computer-controlled solenoid valves triggered flow of individual channels of odorized air (1.0 l min<sup>-1</sup>) diluted in charcoal-deodorized air (1.0 l min<sup>-1</sup>). The odorized air was delivered from the olfactometer to the cue port via Teflon tubing (1.5 mm i.d. × 70 cm). Rats self-initiated each experimental trial by introducing their snout into the cue port. After a random delay of 0.15–0.25 s, either of two odour cues was delivered. The order of odours was altered pseudorandomly across trials. A tone signal was delivered 1.0 s after cue initiation. If the rat stayed in the cue port until the tone signal, food reward was made available in one of the food cups. Odours A, C and E were associated with food in the left cup and odours B, D and F with food in the right cup. Withdrawal of the nose from the cue port terminated the delivery of odour. Rats usually withdrew from the odour port within 1.0–1.5 s and then moved to the cue-associated food cup. If the rat withdrew its nose from the cue port before the tone signal, no food reward was delivered in any of the cups. Correct choices were rewarded with a 45 mg sucrose pellet delivered via a food dispenser (EVN-203-45IR, Med Associates) as soon as the rat broke the photobeam.

**Training procedure.** Animals were first habituated to the behavioural apparatus and the experimenter for one week and then pre-trained for another week to perform nose pokes at the cue port in the absence of odour cues. Nose poking was followed by running to the food cups and consumption of food rewards. After completion of pre-training, odour cues A/B were introduced in a pseudorandom order during the cue sampling interval and the rats were trained to associate one odour with the left food cup and the other with the right cup by being rewarded only for appropriate choices (Fig. 1b). Training was terminated when the rats reached to a criterion of 85% correct rate for two consecutive sessions (Fig. 1c, 16.8 ± 0.8 days). Days 2, 6, 10, 14 and 18 after odour introduction were defined as time-points T1–T5, respectively (Fig. 1c). The animals were trained on at least two sessions of 30 min daily, usually with ~60 trials per session. Surgery was performed after pre-training. For 5 dCA1-LEC rats, associative training sessions were performed after the tetrodes were in place at the regions of interest.

**Non-cued task.** After the animals reached the criterion, they were tested in a non-cued control task, in which only flow of vehicle pure air without odour cues was delivered when the animal made the nose poke. After the poke, the animals made random runs to either of the two food ports, 85% of which were randomly rewarded. Non-cued trials were scheduled after cued sessions each day after the rat had reached the criterion.

**Novel cue task.** In four rats with simultaneous implants in dCA1 and LECs, training was continued with odour cues C/D, using a similar protocol as for A/B. Criterion was reached within 3 days. The same rats were subsequently trained for cues E/F for another 3 days.

**Random foraging task.** On each testing day, after the cue–place association task, the rats were tested in a random foraging task in an open field (100 × 100 cm) to estimate spatial tuning of the cells. The rats were trained for at least one session of 10–20 min. Running was motivated by chocolate cereal crumbs or vanilla cookie crumbs thrown randomly into the enclosure.

**Data collection.** Spike activity was recorded against reference electrodes located at the level of the corpus callosum or higher up for hyperdrive recordings, and against tetrodes with no spike activities in the same region for microdrive recordings. Local field potential (LFP) was always recorded single-ended against a ground in the skull above the cerebellum.

The tetrode drive was connected to a multichannel, impedance matching, unity gain headstage. The output of the headstage was conducted to an Axona acquisition system (Axona Ltd) or to a Neuralynx data acquisition system (Neuralynx) via a lightweight multiwire tether cable and through a slip-ring commutator. Unit activity was amplified by a factor of 3,000–5,000 and band-pass filtered from 800 to 6,700 Hz (Axona) or from 600 to 6,000 Hz (Neuralynx). Spike waveforms above a threshold set by the experimenter (~55 µV) were time-stamped and digitized at 32 kHz (Neuralynx) or 48 kHz (Axona) for 1 ms.

For the hyperdrive recordings in CA1, LFP signals, 1 per tetrode, were recorded in the 0–475-Hz frequency band at a sampling rate of ~2,000 Hz (1,893 Hz sampling and 2,034 Hz sampling for analogue and digital Neuralynx systems, respectively). For the microdrive recordings in dCA1 and LEC, LFP signals from two tetrodes per microdrive were amplified 1,000–3,000 times, lowpass-filtered at 500 Hz, and sampled at a rate of 4,800 Hz (16 bits per sample). This allowed us to analyse the dynamics of oscillatory coherence between dCA1 and LEC for 2 × 2 = 4 'recording pairs' in each animal. In a more restricted analysis, we used only one pair from each animal — the pair that showed the highest power in the 20–40-Hz frequency band (assuming that coherence could be estimated most accurately from this tetrode). Notch filters were not applied. Signals from thermocouples were recorded as described for LFP signals. Light-emitting diodes (LEDs) on the headstage were used to track the rat's movements at a rate of 25 Hz (Neuralynx) or 50 Hz (Axona).



**Data analysis.** Unless indicated otherwise, analyses were performed using MATLAB code written by the authors. Methods for spectral and wavelet analysis were described previously<sup>12</sup>. Analyses were performed for periods of interest (that is, the cue sampling period at the cue port, runs from cue port to food cups, or runs from food cups to the cue port) during individual trials and then averaged across multiple trials within a session. Unless otherwise indicated, data from the cue sampling period were taken from the second half (500–1,000 ms) to reflect the delayed onset of changes in LFP and spike activity (Figs 1e, g, h and 2a, and Extended Data Fig. 9a, e). Data were normalized against the pre-cue period, when the animal was stationary at the odour port before the cue was delivered (150 to 0 ms before cue onset). To be sure that no odour sampling occurred during the baseline period, we included only trials where the animal's snout was outside the odour port for at least 1 s before the trial. In time-resolved representations, data were averaged by setting the timing of odour cue onset as a trigger. For statistical tests, we used parametric tests since the data did not deviate substantially from a normal distribution. *t*-tests were performed when there was similar variance between groups. s.e.m. are provided for all analyses.

**LFP analysis.** Time–frequency analysis of power. Time-resolved power across frequencies was computed using a wavelet transform method. Signals were convolved by a family of complex Morlet's wavelets  $w(t, f)$ , one for each frequency, as a function of time:

$$w(t, f) = A \exp(-t^2/2\sigma_t^2) \exp(2i\pi ft)$$

in which the family of wavelets was characterized by a constant ratio  $f/\sigma_f$ , which was set to 7, and with  $\sigma_f = \frac{1}{2}\pi\sigma_t$ . The coefficient *A* was set at:

$$(\sigma_t \sqrt{\pi})^{-1/2}$$

in order to normalize the wavelets such that their total energy was equal to 1.

To control for impedance differences between tetrodes, LFP power was normalized, for each tetrode, to power during the pre-cue period (150 to 0 ms before cue onset) when the animal was stationary at the odour port before the cue was delivered. Normalized power was averaged across trials that did not include prematurely terminated trials (<1 s) during the 1 s period preceding the onset of cue sampling.

**Coherence.** Coherence was computed using a multitaper method<sup>31</sup> from the Chronux open source MATLAB toolbox (<http://chronux.org/>). Recordings were cut into non-overlapping, 5 s long windows around the cue sampling event. Results for multiple trials in a single session were averaged. We typically used window sizes of 0.4–1 s and 3–5 tapers. To extract periods of high 20–40-Hz coupling, we first computed time-varying coherence within the 20–40-Hz band using 100-ms bins. Bins were collected when coherence exceeded 2 s.d. of the time-averaged coherence.

**Bandpass-filtering.** An acausal (zero phase shift), frequency domain, finite impulse response bandpass filter was applied to the signals. For theta filtering, 5 and 6 Hz were chosen for stopband and passband, respectively, for low cut-off frequencies; 10 and 11 Hz were chosen for passband and stopband high cut-off frequencies. For filtering of 20–40-Hz oscillations, 18 and 20 Hz were chosen for the stopband and passband, respectively, for the low cut-off frequencies; 40 and 42 Hz were chosen for passband and stopband high cut-off frequencies.

**Selection of theta cycles.** Theta cycles were selected by bandpass filtering the signal from 6–10 Hz and selecting local minima in the filtered signal (that is,  $\theta(t-1) > \theta(t)$  and  $\theta(t+1) > \theta(t)$ ). Segments of the recording were collected and defined as a theta cycle if the time between detected points fell within a range criterion that corresponded to the period of an ~8-Hz theta cycle (that is,  $125 \pm 25$  ms). Local minima of detected theta cycles were required to be separated by at least 100 ms.

**Detection of oscillation episodes.** To extract periods of 20–40-Hz oscillatory activity in the LFP, we first computed time-varying power within the frequency bands for each recording. Power at each time point was averaged across the 20–40 Hz frequency range to obtain time-varying estimates of oscillatory power. Time points were collected when the power exceeded 2 s.d. of the time-averaged power. Time windows, 160 ms in length, were cut around the identified time points. In each 160-ms segment, the maxima of 20–40-Hz oscillatory amplitude were determined from the 20–40-Hz bandpass filtered versions of the recordings. Duplicated 20–40-Hz oscillatory periods, a common consequence of extracting overlapping time windows, were avoided by discarding identical maxima values within a given 20–40-Hz oscillatory subtype and further requiring that maxima of a given subtype be separated by at least 100 ms. Individual 20–40-Hz oscillatory windows were finally constructed from the original, non-bandpass filtered recordings as 400 ms long windows centred around the 20–40-Hz oscillatory amplitude maxima.

**Relationship of 20–40-Hz oscillations to theta phase.** LFP recordings were bandpass-filtered in the theta range (6–10 Hz), and theta phases for each time point were estimated using the Hilbert transform function from the Signal Processing Toolbox in MATLAB. Theta phases at the time points associated with 20–40-Hz maxima (determined as described above) as well as theta phases for CA1 place cell spikes were

collected. Theta phases for each 20–40-Hz oscillatory event were sorted into 30° bins, allowing the phase distribution of each event to be determined. For a given recording, the distributions of 20–40-Hz oscillations were normalized by dividing the bins by the total number of 20–40-Hz oscillatory episodes within a given recording. The normalized distributions were averaged across recordings. In this analysis, and in all analyses involving oscillation phase, the oscillation trough was defined as 0°.

**Cross-frequency analysis.** The time-varying power in a particular frequency band of the LFP (from 2–200 Hz, in 2 Hz wide frequency bins) was calculated using a Morlet's wavelets technique (see above). Coherence was estimated between the original signal and the time-varying power of the signal for each frequency of interest using the magnitude squared coherence function (Hanning window with 50% overlap; 16,384 and 4,096 FFT points for Axona and Neuralynx data, respectively) from the Signal Processing Toolbox in MATLAB. This provided a measure for determining if power changes at particular frequencies (for example, 20–40 Hz) were correlated with phase at other frequencies (for example, theta).

**Time–frequency representation of power across individual theta cycles.** Time-varying power in 2-Hz-wide frequency bands, from 2 to 100 Hz, was obtained for individual theta cycles using the wavelet transform method described above. Time frequency representations for multiple theta cycles recorded from the same site and session within the same animal were then averaged.

**Prewhitening.** For illustrations of power shown in Fig. 2e, a prewhitening autoregressive filter (order of 5) was applied before computation of time frequency analysis to de-emphasize the lower frequency end of the spectrum for illustration purposes. Filter coefficients were estimated using a Yule-Walker approach. Prewhitening was not applied in any of the analyses; it was only used to better visualize power at high frequencies in the selected example above.

**Speed–frequency correlation.** The relationship between running speed and LFP frequency was estimated following procedures described previously<sup>19</sup>. Briefly, we first calculated the absolute power spectrum of LFP during each 200 ms time window. Relative power was then calculated by determining a *z*-score for each 0.49 Hz frequency bin across successive time windows of individual sessions. Finally, speed was correlated with peak frequency for each speed bin.

**Sniffing analysis.** Sniffing was measured by thermocouples implanted in the right nostril of the animals<sup>32</sup>. In each breathing cycle, inhalation is associated with decaying voltage output from the thermocouple, and exhalation with increasing voltage output. Signals from the thermocouples were bandpass-filtered at 0.5 and 1.5 Hz for stopband and passband, respectively, for low cut-off frequencies, and at 44.5 and 45.5 Hz for passband and stopband for high cut-off frequencies. Troughs of the signals were detected and the duration of individual sniffs was determined. The inverse of individual sniffing periods was taken as the instantaneous frequency of the sniffing.

**Spike analysis.** Spike sorting and cell classification. Spike sorting was performed offline using graphical cluster-cutting software described previously<sup>29,33</sup>. Putative excitatory cells were distinguished from putative interneurons by spike width and average rate. Firing rate maps (see below) and head direction tuning curves were used for classification of cell types in MEC<sup>33–36</sup>. Putative excitatory cells with mean firing rates of more than 1.0 Hz (mean firing rate during all left and right cue sampling intervals) were included for further analysis.

Since the tetrodes were not moved during task acquisition, most of the units included in the recordings at time points T1–T5 were from the same cell population. However, occasionally some cells were lost during the course of learning whereas other cells were added, causing slight variation in the number of the cells from session to session (from T1 to T5: between 63 and 75 cells in dCA1 and between 76 and 82 cells in LEC). Cells on error trials (T5e) were from the same population as recorded on correct T5 trials and on down-sampled T5 trials (T5d).

**Phase-locking of spikes to 20–40-Hz oscillations.** Spikes that occurred during 20–40-Hz oscillatory episodes were selected. For these oscillation-associated spikes, oscillatory phase at the time of spike occurrence was estimated. This was performed by first bandpass-filtering the LFP, performing a Hilbert transform on the filtered signal, and then extracting the phase component at the spike times. Cells were considered to be phase-locked if their phase distribution differed significantly from a uniform distribution ( $P < 0.05$ , Rayleigh test).

**Selectivity index.** Odour-specific representation of spikes was assessed using a selectivity metric. Selectivity was computed by comparing firing rates during sampling for left-associated and right-associated cues on successive 100-ms bins:

$$\text{Selectivity} = (FR_1 - FR_2) / (FR_1 + FR_2)$$

where  $FR_1$  and  $FR_2$  are mean firing rates across multiple trials in individual 100-ms bins during left- and right-associated cue samplings, respectively.

To confirm the results with this difference-over-sum index, we also computed Selectivity using receiver-operating characteristic (ROC) analysis<sup>37</sup>, a method based on signal detection theory:

$$\text{Selectivity}_{\text{ROC}} = 2 \times (\text{auROC} - 0.5)$$

where auROC is the area under the ROC curve computed from spike numbers for left and right trials in individual 100-ms bins. auROC = 0.5 indicates that the firing rates were not different between left and right trials; auROC = 1 indicates that firing rates in all left-cue sampling trials were larger than all right-cue sampling trials (see Extended Data Fig. 10).

After scaling, for both Selectivity metrics, Selectivity = 1 indicates that all firing occurs during the left-associated cue odour, whereas Selectivity = -1 means that all firing occurs during right-cue odours.

**Population vector analysis.** To estimate the development of odour-specific representations for a whole population of cells, population vector (PV) analysis was performed as described previously<sup>38,39</sup>. Mean firing rates in individual 100-ms bins were pooled for cells from all animals to construct vectors for left and right-associated odours. A correlation coefficient (CC) for the two vectors was computed and then re-scaled as a PV Differential Index (PDI) to emphasize the difference between the population vectors as

$$\text{PDI} = 1 - \text{CC}$$

PDI = 0 indicates that the population vectors were identical, whereas PDI = 1 denotes that activity was maximally different between left and right-associated trials.

The chance level of the PDI was determined by a random permutation procedure using all cells recorded at that time-point in each region. One thousand permutations were performed for each cell in the sample. For each permutation trial, individual cue-sampling trials were randomly re-assigned as left or right-associated trials, maintaining the number of left and right trials identical to the original data. Population vectors were computed from the same ensemble of the cells as the original data for each of 1,000 permutations, resulting in 1,000 sets of PDI. The distribution of PDI across all 1,000 permutations in the sample was computed and finally the 95th percentile was determined as the chance level.

**Firing rate maps.** Position estimates were based on tracking of the LEDs on the head stage connected to the tetrode drive. For rate maps in the random foraging task, data were speed-filtered; only epochs with instantaneous running speeds of  $5 \text{ cm s}^{-1}$  or more were included.

To characterize firing fields, the position data were sorted into  $2.5 \text{ cm} \times 2.5 \text{ cm}$  bins. The path was smoothed with a 21-sample boxcar window filter (400 ms; 10 samples on each side). Firing rate distributions were then determined by counting the number of spikes in each bin as well as the time spent per bin. Maps for number of spikes and time were smoothed individually using a boxcar average over the surrounding  $5 \times 5$  bins. Weights were distributed as follows:

$$\text{boxcar weights} = \begin{bmatrix} 0.0025 & 0.0125 & 0.0200 & 0.0125 & 0.0025 \\ 0.0125 & 0.0625 & 0.1000 & 0.0625 & 0.0125 \\ 0.0200 & 0.1000 & 0.1600 & 0.1000 & 0.0200 \\ 0.0125 & 0.0625 & 0.1000 & 0.0625 & 0.0125 \\ 0.0025 & 0.0125 & 0.0200 & 0.0125 & 0.0025 \end{bmatrix}$$

**Spatial information.** For each cell, the spatial information content in bits per spike was calculated as

$$\text{information content} = \sum_i p_i \frac{\lambda_i}{\lambda} \log_2 \frac{\lambda_i}{\lambda}$$

where  $\lambda_i$  is the mean firing rate of a unit in the  $i$ th bin,  $\lambda$  is the overall mean firing rate, and  $p_i$  is the probability of the animal being in the  $i$ th bin (occupancy in the  $i$ th bin / total recording time)<sup>40</sup>. An adaptive smoothing method, introduced by Skaggs *et al.*<sup>41</sup>, was used before the calculation of information scores<sup>29</sup>.

Place cells were defined as cells with spatial information content above the chance level for this parameter<sup>42</sup>. The chance level was determined for each time-point during learning in each brain region by a random permutation procedure using all cells recorded at that time-point in that region. One hundred permutations were performed for each cell in the sample. For each permutation trial, the entire sequence of spikes fired by the cell was time-shifted along the animal's path by a random interval between 20 s and 20 s less than the total length of the trial (usually 600 – 20 = 580 s), with the end of the trial wrapped to the beginning to allow for circular displacements. This procedure allowed the temporal firing structure to be retained in the shuffled data at the same time as the spatial structure was lost. Spatial information was then calculated for each shuffled map. The distribution of spatial information values across all 100 permutations of all cells in the sample was computed and finally the 95th percentile was determined. Place cells were defined as cells with spatial information scores above the 95th percentile of the distribution from shuffled data for the relevant group.

**Gridness score.** The structure of all rate maps was evaluated for all cells by calculating the spatial autocorrelation for each smoothed rate map<sup>33,42,43</sup>. The degree

of spatial periodicity ('gridness score') was determined for each recorded cell by taking a circular sample of the autocorrelogram, centred on the central peak but with the central peak excluded, and comparing rotated versions of this sample<sup>42</sup>. Grid cells were defined as cells in which rotational symmetry-based grid scores exceeded the 95th percentile of a distribution of grid scores for shuffled recordings from the entire population of cells. Shuffling was performed in the same way as for place cells by time-shifting the spike sequence for each permutation trial. The shuffling procedure was repeated 100 times for each of the cells in the sample. Grid scores were computed for all permutations for all cells and the 95th percentile was read out from the overall distribution. Grid cells were defined as cells with grid scores higher than the 95th percentile of the grid scores of the distribution for the shuffled data.

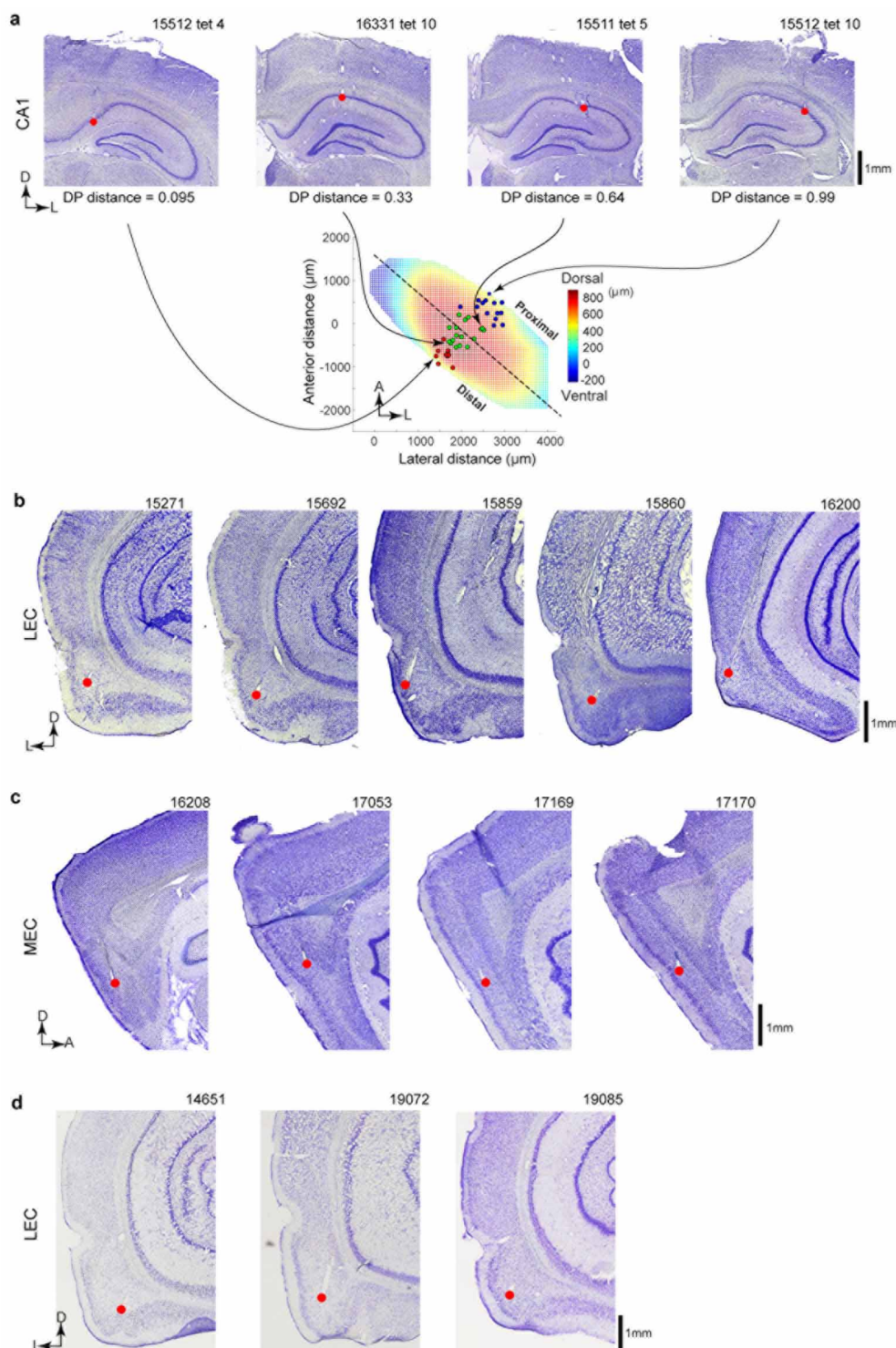
**Histology and 3D reconstruction.** After receiving an overdose of pentobarbital, the rats were perfused intracardially with saline and 4% formaldehyde. The brains were then removed and stored in formaldehyde. For verification of tetrode locations, the brains were frozen, cut at 30  $\mu\text{m}$  thickness, and stained with cresyl violet. For hyperdrive recordings, all tetrodes of the 14-tetrode bundle were identified, and the tip of each electrode was found by comparison with adjacent sections. For microdrive recordings, the positions of the bundle of four tetrode tips were determined. Digital photomicrographs were acquired with a Zeiss Axioimager-Z1 microscope equipped with a digital camera using AxioVision software. For CA1 recordings, all tetrodes were identified in the pyramidal cell layer of the CA1. For LEC and MEC recordings, all tetrodes were identified in layer III. In some animals with LEC or MEC microdrives, LFP recordings were moved for 100  $\mu\text{m}$  to acquire additional cells after completion of T5 recordings with the original tetrode locations. In these cases, the exact position of the electrodes at the time of the original recording was extrapolated<sup>42</sup>.

For CA1 recordings, the positions of the tetrodes were three-dimensionally reconstructed and analysed using as described previously<sup>44</sup>. Tetrode positions and structure of cell layers for each section in the dorsal hippocampal region were digitally registered using a computer-assisted system (NeuroLucida, MBF Bioscience) according to the manufacturer's instructions. The tracing output from NeuroLucida was imported into MATLAB (Mathworks) for further analyses. To place electrode positions in identical coordinates and minimize the distortions originating from sectioning and reconstruction, reconstructed neurons were linearly transformed and fitted into reference brain coordinates. A reference brain was first reconstructed. Three landmarks for transformation were used: the anterior tip of subiculum (tSUB), the anterior tip of the lateral ventricle, and the 'X'-shaped pyramidal pool of CA3 in the intermediate portion of the hippocampus<sup>45</sup>. These landmarks were selected because (1) they were distributed at the periphery of the dorsal hippocampus and (2) they could be uniquely defined and were unaffected by rotation of the brain. The coordinate system was defined as giving the tSUB coordinates of [0, 0, 0] in the [x (lateral), y (anterior) and z (dorsal)] directions (in  $\mu\text{m}$ ). Individual sample brains with the tetrode tip positions were then fitted to the reference brain with regard to the above landmarks using affine (linear) transformation (translation, rotation, scaling, and shear). The surface of CA1 pyramidal cell layer was then reconstructed by interpolating reconstructed cell layer data. The major (longest) axis of the CA1 cell layer was defined as the septotemporal axis of the hippocampus. For each tetrode tip, distal-proximal (DP) distance was calculated three-dimensionally by the following procedure: a transverse plane that includes the tetrode position and is perpendicular to the septotemporal axis is determined. An intersectional curve between this transverse plane and the CA1 surface is then calculated. DP distance  $D_{\text{DP}}$  was defined as  $D_{\text{DP}} = L_{\text{DT}}/L_{\text{DP}}$ , where  $L_{\text{DT}}$  was the length of the curve between the distal ending point of the intersectional curve and the tetrode tip position, and  $L_{\text{DP}}$  was the length between the distal and proximal two ending points along the curve.  $D_{\text{DP}} = 0$  indicates the distal border, whereas  $D_{\text{DP}} = 1$  indicates the proximal border.

29. Henriksen, E. J. *et al.* Spatial representation along the proximodistal axis of CA1. *Neuron* **68**, 127–137 (2010).
30. Steward, O. Topographic organization of the projections from the entorhinal area to the hippocampal formation of the rat. *J. Comp. Neurol.* **167**, 285–314 (1976).
31. Mitra, P. P. & Pesaran, B. Analysis of dynamic brain imaging data. *Biophys. J.* **76**, 691–708 (1999).
32. Kepecs, A., Uchida, N. & Mainen, Z. F. Rapid and precise control of sniffing during olfactory discrimination in rats. *J. Neurophysiol.* **98**, 205–213 (2007).
33. Sargolini, F. *et al.* Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* **312**, 758–762 (2006).
34. Solstad, T., Boccara, C. N., Kropff, E., Moser, M. B. & Moser, E. I. Representation of geometric borders in the entorhinal cortex. *Science* **322**, 1865–1868 (2008).
35. Hafting, T., Fyhn, M., Molden, S., Moser, M. B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
36. Fyhn, M., Molden, S., Witter, M. P., Moser, E. I. & Moser, M. B. Spatial representation in the entorhinal cortex. *Science* **305**, 1258–1264 (2004).
37. Macmillan, N. A. & Creelman, C. D. *Detection Theory: A User's Guide* (Cambridge Univ. Press, 1991).

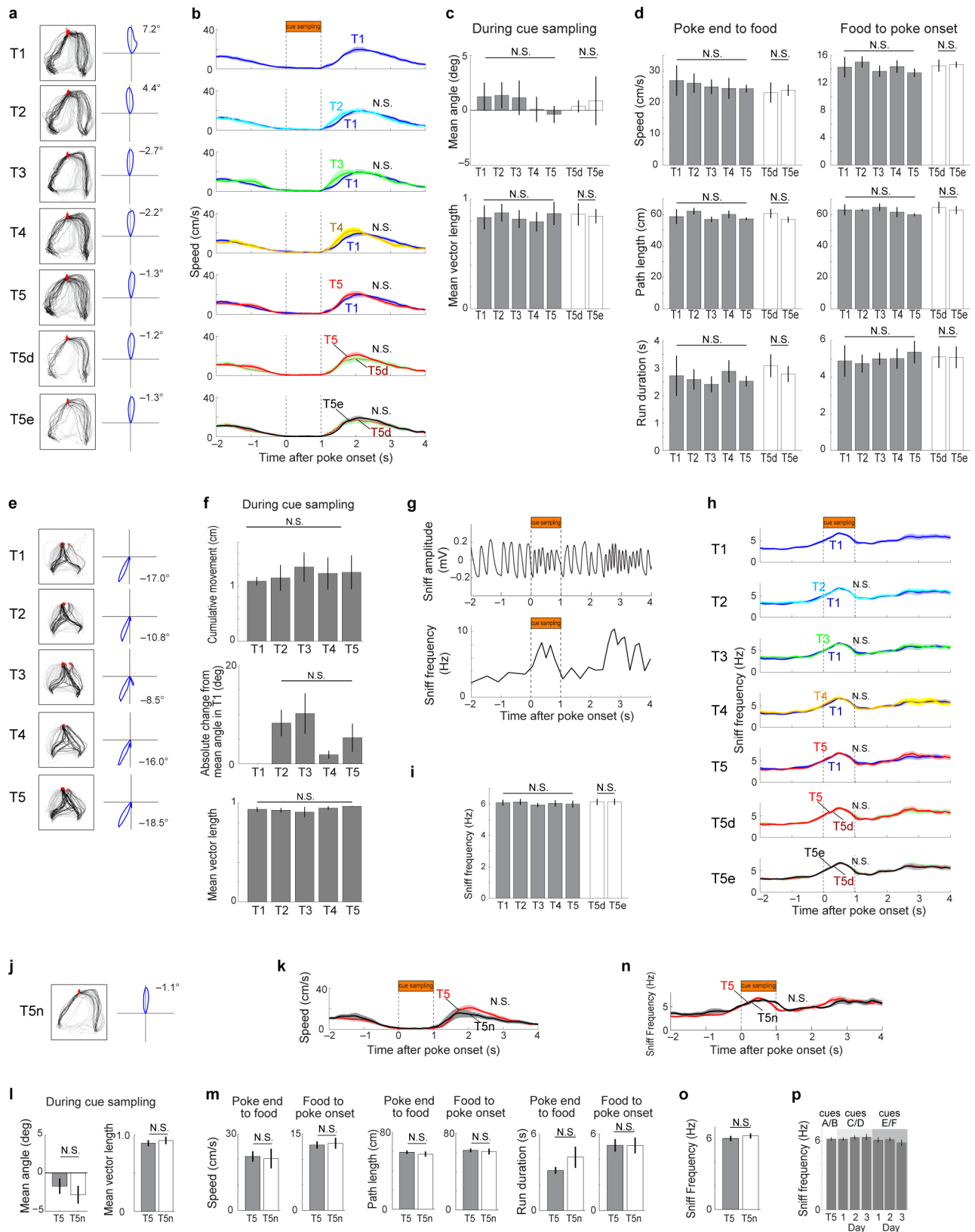
38. Jezek, K., Henriksen, E. J., Treves, A., Moser, E. I. & Moser, M. B. Theta-paced flickering between place-cell maps in the hippocampus. *Nature* **478**, 246–249 (2011).
39. Naya, Y. & Suzuki, W. A. Integrating what and when across the primate medial temporal lobe. *Science* **333**, 773–776 (2011).
40. Skaggs, W. E., McNaughton, B. L., Gothard, K. M. & Markus, E. J. in *Advances in Neural Processing Systems* Vol. 5 (eds Hanson, S. J., Cowan, J. D. & Giles, C. L.) 1030–1037 (Morgan Kaufmann, San Mateo, 1993).
41. Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
42. Langston, R. F. *et al.* Development of the spatial representation system in the rat. *Science* **328**, 1576–1580 (2010).
43. Fyhn, M., Hafting, T., Treves, A., Moser, M. B. & Moser, E. I. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* **446**, 190–194 (2007).
44. Igarashi, K. M. *et al.* Parallel mitral and tufted cell pathways route distinct odor information to different targets in the olfactory cortex. *J. Neurosci.* **32**, 7970–7985 (2012).
45. Fanselow, M. S. & Dong, H. W. Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron* **65**, 7–19 (2010).





**Extended Data Figure 1 | Brain sections showing recording positions in CA1, LEC and MEC.** **a**, Recording positions (red) in CA1 in coronal brain sections from representative animals (top; animal and tetrode numbers indicated; distal-proximal (DP) distances as in Extended Data Fig. 5) and in a projection of a 3D reconstruction of the CA1 pyramidal cell layer (bottom; see Extended Data Fig. 5 for the full 3D reconstruction). Abscissa shows the lateral-medial axis and ordinate the anterior-posterior axis. Dorsal-ventral is colour-coded. Each dot corresponds to one recording location in distal (red), intermediate (green) and proximal (blue) parts of CA1. Stippled line indicates

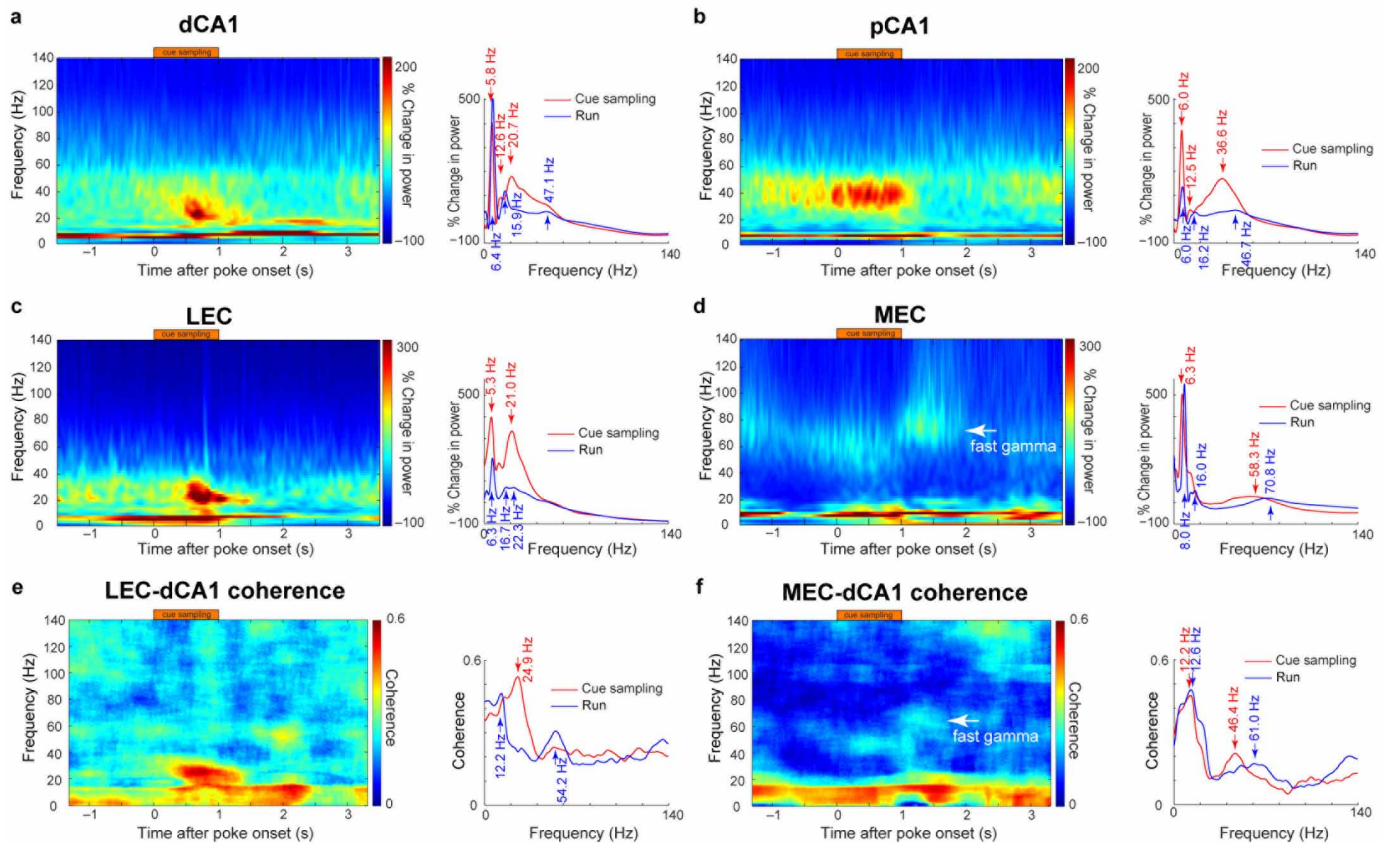
septotemporal axis of the hippocampus (see Supplementary Methods). **b**, Tetrode positions (red) in coronal sections through LEC for all 5 animals with simultaneous recordings in LEC and dCA1. Animal numbers are indicated. **c**, Tetrode positions (red) in sagittal sections through MEC (all 4 animals with simultaneous MEC and dCA1 recordings). **d**, Tetrode positions (red) in coronal brain sections through LEC from all 3 animals with simultaneous recordings in pCA1 and LEC. Orientation: D, dorsal; L, lateral; A, anterior.



**Extended Data Figure 2 | Learning was not associated with changes in motor behaviours.** **a**, Left, example trajectories of one animal on sessions T1 to T5, on error trials (T5e) and on correct trials downsampled to the same number of trials as the error trials (T5d). Red parts of each trajectory indicate positions covered during the cue sampling interval, black indicates positions during runs from cue port to food cups, and grey shows positions from food cups back to the cue port. Right, polar plots showing distributions of head direction during the cue sampling interval (T1–T5, T5d and T5e, as indicated to the left). Values to the right indicate mean head angle relative to the centre of the cue port, designated as north ( $0^\circ$ ). **b**, Time course of instantaneous speed before, during and after cue sampling, averaged across 5 rats. Shading denotes s.e.m. T1 is shown at the top; T5, T5d and T5e at the bottom. For every 10-ms time bin during T2–T5, speed was compared with the corresponding bin at T1. Speeds were not significantly different from T1 at any time bin ( $q > 0.05$ ; false discovery rate (FDR) corrected for multiple comparisons). Speed was not different at T5e and T5d. **c**, Mean head angle (top) and mean vector length for head direction (bottom) during the cue sampling interval (mean  $\pm$  s.e.m. for all 5 animals). Neither head angle nor mean vector length changed significantly from T1 to T5 (repeated measures ANOVA:  $F(4, 16) = 0.44$ ,  $P = 0.78$  for mean angle;  $F(4, 16) = 0.13$ ,  $P = 0.97$  for mean vector length). There was also no change in these parameters on the error trials (T5e compared with T5d using two-tailed paired  $t$ -test:  $t(4) = 0.70$ ,  $P = 0.52$  for mean angle;  $t(4) = 0.18$ ,  $P = 0.87$  for mean vector length). **d**, Mean instantaneous speed (top), path length (middle) and run duration (bottom) for trajectories from cue port to food cups (left column) and from food cups to cue port (right column). None of these parameters changed significantly during learning (repeated measures ANOVA for time-points T1–T5; for cue port to food cups:  $F(4, 16) = 0.14$ ,  $P = 0.97$  for mean speed,  $F(4, 16) = 0.55$ ,  $P = 0.70$  for path length,  $F(4, 16) = 0.16$ ,  $P = 0.95$  for run duration; for food cups to cue port:  $F(4, 16) = 0.35$ ,  $P = 0.84$  for mean speed,  $F(4, 16) = 0.47$ ,  $P = 0.76$  for path length,  $F(4, 16) = 0.13$ ,  $P = 0.97$  for run duration). None of these parameters changed on error trials (T5e compared with T5d using two-tailed paired  $t$ -test; cue port to food cups:  $t(4) = 0.31$ ,  $P = 0.77$  for mean speed,  $t(4) = 1.87$ ,  $P = 0.52$  for path length,  $t(4) = 0.91$ ,  $P = 0.42$  for run duration; food cups to cue port:  $t(4) = 0.14$ ,  $P = 0.89$  for mean speed,  $t(4) = 0.50$ ,  $P = 0.64$  for path length,  $t(4) = 0.02$ ,  $P = 0.99$  for run duration). **e**, To check if changes in neural activity from T1 to T5 are associated with changes in body position, LEDs were attached to the back of 4 rats and LED positions were recorded from T1 to T5. Left, example trajectories of body position in one animal at different stages of learning. Red indicates body positions covered during the cue sampling interval, black indicates positions during runs from cue port to food cup, and grey show positions from food cup back to the cue port. Right, polar plots showing distribution of body direction (deflection from south) during the cue sampling interval (T1–T5). **f**, Cumulative body movement (top), mean body angle (middle) and mean vector length for body direction (bottom) during the cue sampling interval (mean  $\pm$  s.e.m. of 4 animals). In the middle panel, body angle is shown as the absolute change in mean angle compared to T1, since different rats may turn in different directions. None of these parameters changed significantly from T1 to T5 (repeated measures ANOVA:  $F(4, 12) = 0.77$ ,  $P = 0.57$  for cumulative body movement;  $F(3, 9) = 1.06$ ,  $P = 0.41$  for mean angle;  $F(4, 12) = 1.11$ ,  $P = 0.40$  for mean vector length). **g**, Example recording

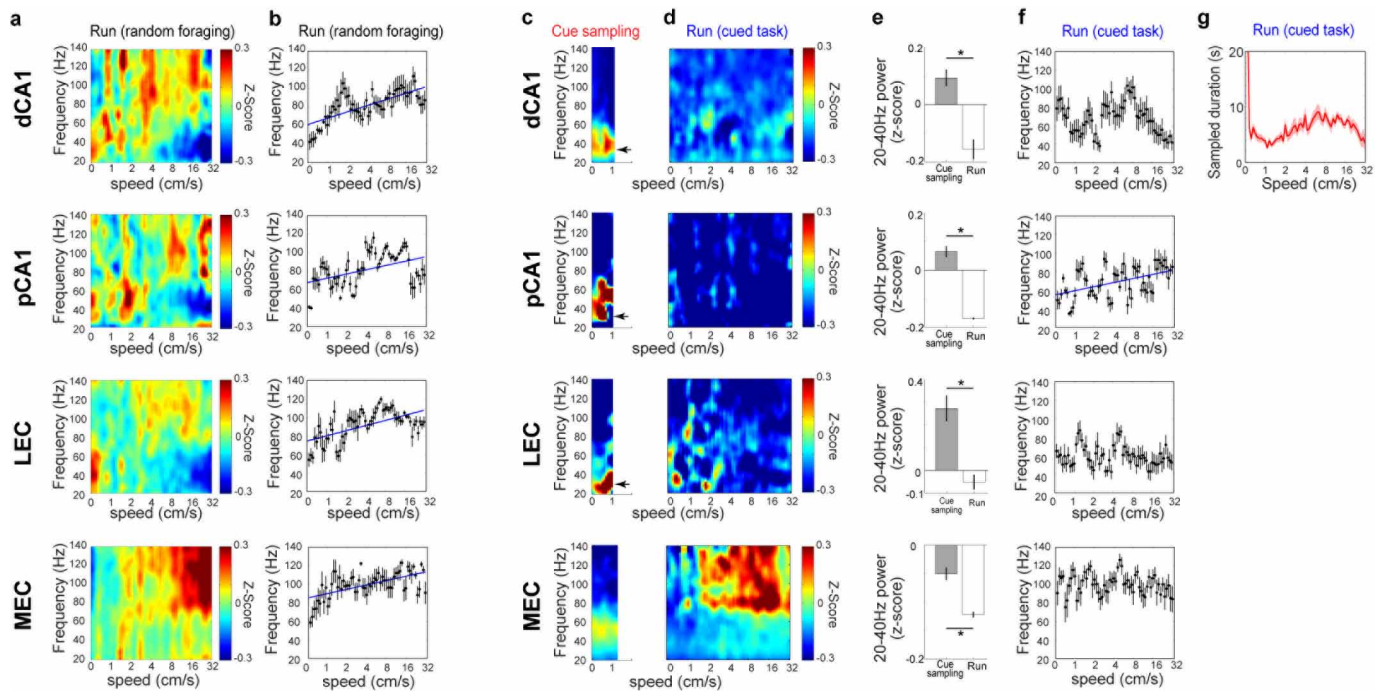
showing changes in sniffing amplitude and frequency during cue sampling. A temperature sensor (thermocouple) was implanted in the right nostril to measure respiration (top). In each breathing cycle, inhalation is associated with decaying voltage output from the thermocouple, and exhalation with increasing voltage. Instantaneous sniff frequency was determined from the voltage change (bottom). Note that sniff frequency increased during cue sampling (between  $t = 0$  s and  $t = 1$  s) and when the animal approached the food cup. The food cup was reached at  $t = 3.8$  s in the present example. **h**, Time course of sniff frequency averaged for 5 rats at time-points T1–T5 as well as on error trials (T5e) and correct trials down-sampled to the same number of trials (T5d). Shading denotes s.e.m. At time points T2–T5, sniff frequency was compared for every 10 ms bin with corresponding bins at T1. None of the comparisons were significantly different ( $q > 0.05$ ; FDR corrected for multiple comparisons). Sniff frequencies between T5d and T5e were also not different. **i**, Mean sniff frequency during the cue sampling interval. Sniff frequency did not change during the course of learning (repeated measures ANOVA:  $F(4, 16) = 0.24$ ,  $P = 0.91$ ) or on error trials (T5e compared with T5d using two-tailed paired  $t$ -test,  $t(4) = 0.03$ ,  $P = 0.97$ ). **j**, Left, example traces showing position of the animal on non-cued control trials (T5n). Colours indicate positions during cue sampling (red), positions from cue port to food cups (black) and positions from food cups back to cue port (grey). Right, polar plot showing head direction distribution during the cue sampling period for the session shown on the left. Values on the right indicate mean head angle. The centre of the cue port is north, or  $0^\circ$ . **k**, Instantaneous speed on non-cued trials, averaged across 5 rats. Shading denotes s.e.m. Speed on non-cued trials (T5n) was compared at successive 10-ms bins with speed on corresponding bins of cued trials (T5). Differences were not significant at any time bin between 2 s before and 4 s after poke onset ( $q > 0.05$ , FDR corrected for multiple comparisons). **l**, Mean head angle (top) and mean vector length of head direction (bottom) during the cue-sampling interval. Neither head angle nor mean vector length was different from corresponding values on cued trials (two-tailed paired  $t$ -test,  $t(4) = 1.85$ ,  $P = 0.13$  for mean angle;  $t(4) = 1.99$ ,  $P = 0.12$  for mean vector length). **m**, Mean speed (top), path length (middle) and run duration (bottom) for trajectories from cue port to food cups (left column) and from food cups to cue port (right column). None of these parameters were different between cued and non-cued tasks (two-tailed paired  $t$ -test; cue port to food cup movements:  $t(4) = 0.16$ ,  $P = 0.88$  for mean speed,  $t(4) = 0.50$ ,  $P = 0.64$  for path length,  $t(4) = 1.03$ ,  $P = 0.36$  for run duration; food cup to cue port movements:  $t(4) = 0.21$ ,  $P = 0.84$  for mean speed,  $t(4) = 0.30$ ,  $P = 0.78$  for path length,  $t(4) = 0.44$ ,  $P = 0.68$  for run duration). **n**, Sniff frequency at the cue port on non-cued trials, averaged across 5 rats. Frequency on cued trials is shown as a reference. Shading denotes s.e.m. Sniff frequency was compared at successive 10-ms bins between T5 and T5n. Differences were not significant at any time between 2 s before and 4 s after poke onset ( $q > 0.05$ ; FDR corrected for multiple comparisons). **o**, Mean sniff frequency during the cue sampling period was not different between T5 and T5n (two-tailed paired  $t$ -test,  $t(4) = 0.89$ ,  $P = 0.43$ ). **p**, Mean sniff frequency during the cue sampling interval in novel odour trials. Sniff frequency did not change over the course of trials with novel odours (repeated measures ANOVA:  $F(6, 18) = 0.77$ ,  $P = 0.60$ ).





**Extended Data Figure 3 | Power and coherence over a broader spectrum of frequencies in dCA1 and LEC at the end of training.** **a–d**, Time-resolved power spectra averaged across tetrodes in **a**, dCA1 ( $n = 9$  tetrodes), **b**, pCA1 ( $n = 14$  tetrodes), **c**, LEC ( $n = 10$  tetrodes) and **d**, MEC ( $n = 8$  tetrodes), as in Fig. 1e and g, but for a broader band of the frequency spectrum (0–140 Hz). Power is shown as percentage change from power during the pre-cue period. dCA1 and pCA1 data are from the same animals (rats with tetrodes along the entire proximodistal CA1 axis), whereas LEC and MEC data are from separate animals (5 and 4 rats, respectively). Right panels show mean power for frequencies up to 140 Hz during cue sampling (red) and during running from cue port to food cups (blue). Peak frequencies are indicated. In CA1, the mean power in the 20–40-Hz frequency band increased from  $(1.07 \pm 0.04) \times 10^{-3}$

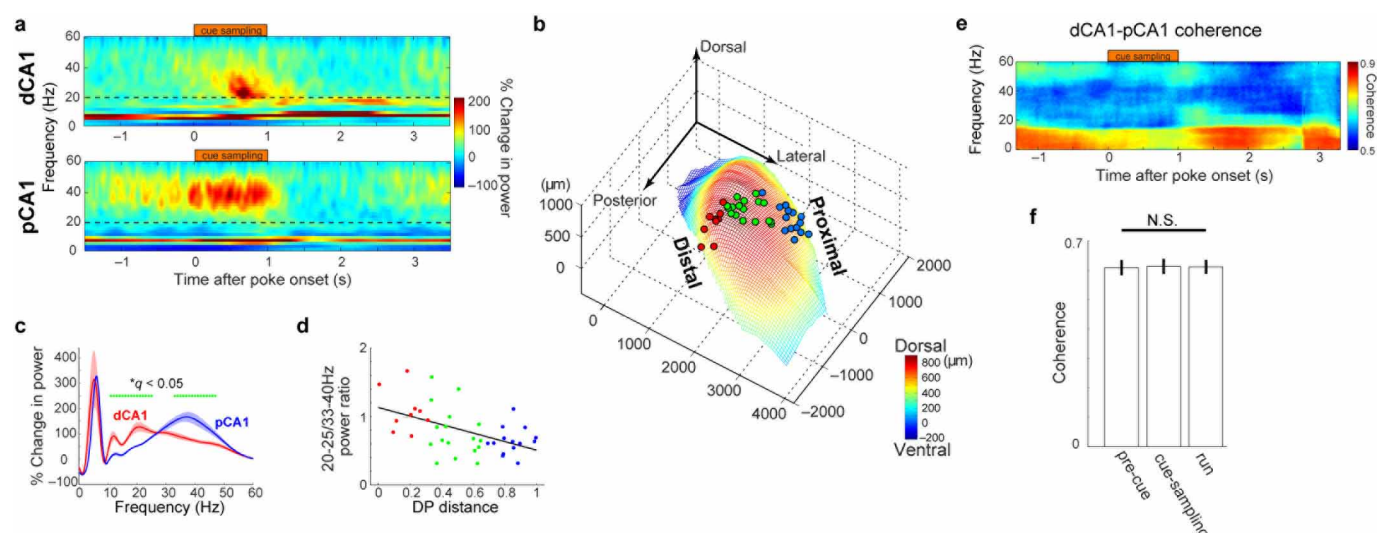
before cue sampling to  $(1.36 \pm 0.05) \times 10^{-3}$  during cue sampling and then reverted to  $(0.89 \pm 0.05) \times 10^{-3} \text{ mV}^2$  after cue sampling (repeated measures ANOVA:  $F(2, 78) = 118, P < 0.001$ ). In LEC, the power was  $(1.05 \pm 0.12) \times 10^{-3}$ ,  $(1.62 \pm 0.16) \times 10^{-3}$  and  $(0.78 \pm 0.08) \times 10^{-3} \text{ mV}^2$ , respectively ( $F(2, 18) = 67.1, P < 0.001$ ). **e, f**, Time-resolved coherence spectra averaged across tetrode pairs in LEC and dCA1 (**e**) and in MEC and dCA1 (**f**), as shown in Fig. 1h, but across a broader band of the frequency spectrum (same animals as in **c** and **d**; data are averaged across all EC–dCA1 tetrode pairs). Note that, during running, MEC shows fast gamma oscillations (60–100 Hz) that are coherent with dCA1 LFPs (arrows). Right panels show mean coherence spectra during cue sampling (red) and during running (blue), with peak frequencies indicated. LEC did not show fast gamma oscillations.



#### Extended Data Figure 4 | Effect of running speed on oscillation frequency.

**a**, Power spectra of oscillations in dCA1, pCA1, LEC and MEC as a function of running speed at T5 during random foraging in the  $1 \times 1$  m square box. Power is normalized at each frequency bin and shown as a z-score<sup>18</sup>. Data were averaged across 9 (dCA1), 14 (pCA1), 10 (LEC) and 8 (MEC) tetrodes from 5 dCA1 and pCA1, 5 LEC and 4 MEC rats, respectively. **b**, Peak oscillation frequency (mean  $\pm$  s.e.m. for all tetrodes) as a function of speed bin in **a**. As in previous work<sup>18</sup>, a significant correlation was observed between speed and peak frequency of oscillations in the 20–140-Hz range ( $P < 0.001$ ;  $r(58) = 0.71, 0.44, 0.60$  and  $0.59$  for dCA1, pCA1, LEC and MEC, respectively). **c, d**, Speed–frequency relationship in the cue–place association task. Same tetrodes as in **a**. Data were plotted during the cue sampling period (**c**) and during subsequent running from the cue sampling port to the food cups (**d**). Note that during cue

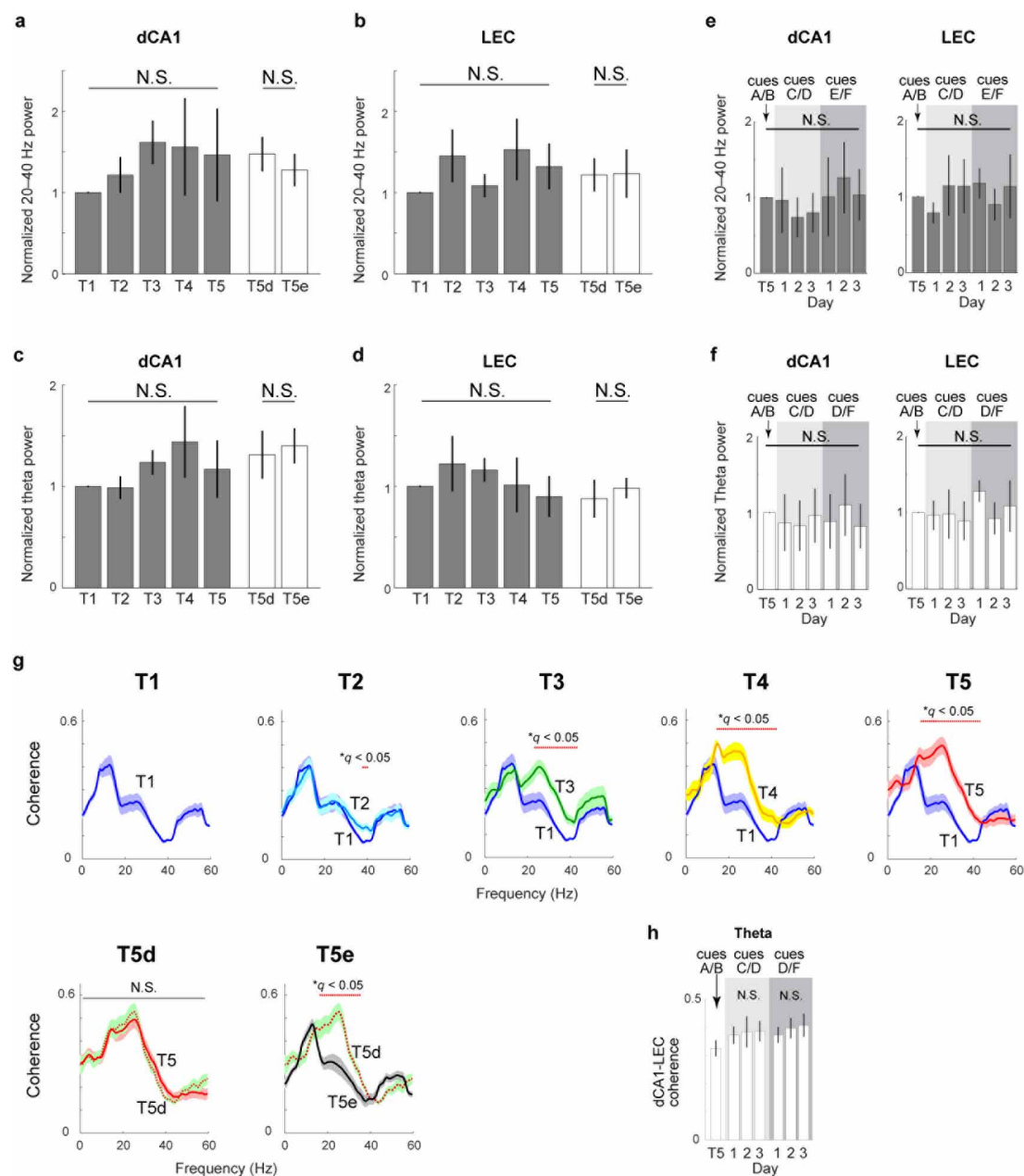
sampling, speed was minimal ( $\sim 0$ – $1$  cm s<sup>-1</sup>). Strong 20–40-Hz oscillations were observed in dCA1, pCA1 and LEC (arrows in **c**) but not at similar speeds during running (**d**). **e**, In all four brain regions (dCA1, pCA1, LEC and MEC), the power at very low speeds ( $0$ – $1$  cm s<sup>-1</sup>) was significantly stronger during cue sampling than during running (paired *t*-test,  $P < 0.01$ ;  $t(8) = 3.9$ ,  $t(13) = 13.7$ ,  $t(9) = 13.4$  and  $t(7) = 4.5$  for dCA1, pCA1, LEC and MEC, respectively), suggesting 20–40-Hz oscillations do not reflect low speed as such. **f**, During running from cue port to food cups, a significant positive correlation between speed and frequency was observed only in pCA1 ( $P < 0.001$ ,  $r(58) = 0.48$ ). **g**, Time spent in each speed bin during the running part of the cued task (means  $\pm$  s.e.m. for all 14 rats). Note that all speed bins (including low speeds) were sampled for 2.5 s or more.



**Extended Data Figure 5 | 20–40-Hz oscillations in dCA1 and pCA1 are functionally decoupled.** **a**, Time-resolved power spectra averaged across all tetrodes in the distal one-third of CA1 (dCA1, top) as well as the proximal one-third (pCA1, bottom) (9 and 14 recording sites, respectively). Data were pooled over 5 rats with hyperdrive implants spanning a wide transverse range of CA1. To control for impedance differences between tetrodes, LFP power during cue sampling was normalized, for each tetrode, to power during the pre-cue period when the animal was stationary at the odour port before the cue was delivered. Normalized power was averaged across tetrodes. Time  $t = 0$  indicates cue onset. **b**, Positions of all tetrodes from which data were recorded in **a**, plotted in a 3D reconstruction of the CA1 cell layer. Distance from the anterior tip of subiculum is shown in micrometres. Position on the distal–proximal (DP) axis (0, distal; 1, proximal; see Supplementary Methods) was calculated in 3D space for each tetrode. CA1 tetrodes were grouped into distal (red point), intermediate (green point) and proximal (blue point) groups, each corresponding to one-third of the DP axis. **c**, Power spectra of LFP averaged over tetrodes in distal and proximal one-thirds of CA1 during cue sampling.

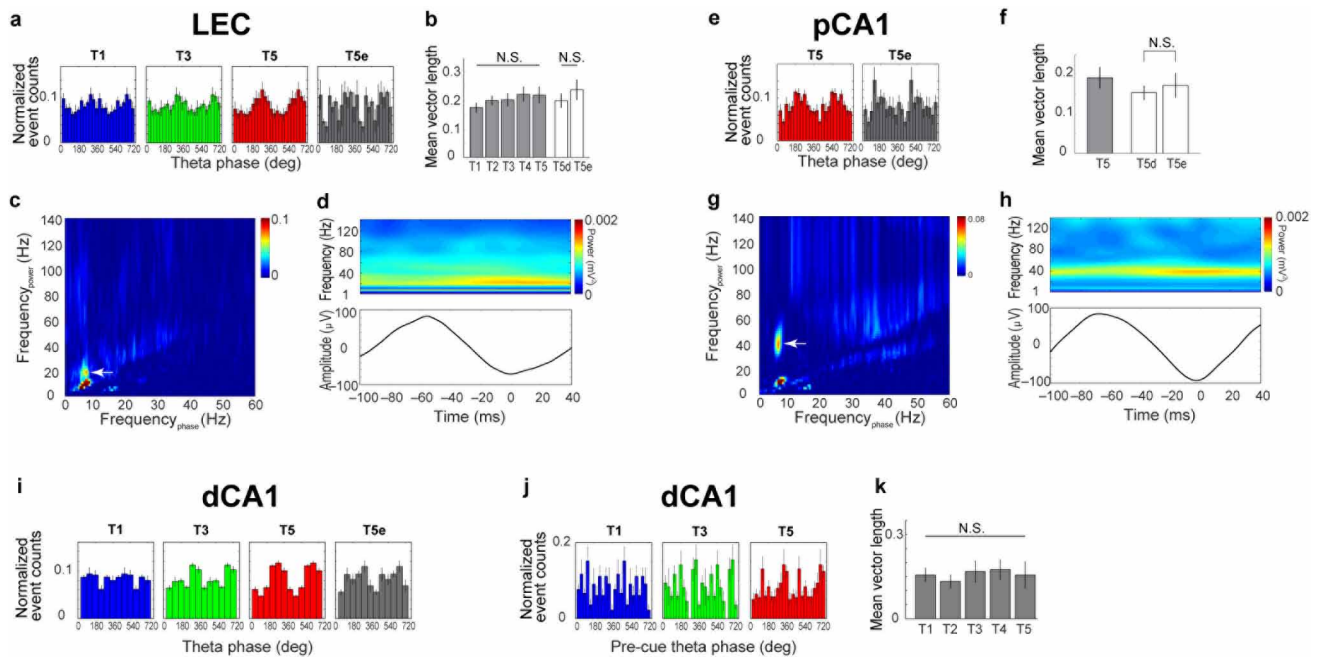
Power is shown as percentage change from the pre-cue level. During cue sampling, power in the lower part of the 20–40-Hz band was stronger in dCA1 than pCA1 (11–25 Hz, green dots;  $q < 0.05$ ; false discovery rate (FDR) corrected for 60 multiple comparisons at 1–60 Hz using 1 Hz bins,  $q < 0.05$ ), whereas power in the higher part was lower in dCA1 (33–47 Hz, green dots;  $q < 0.05$ ). **d**, To compare power in the lower and higher parts of the 20–40-Hz oscillation across the DP axis in CA1, the ratio of power in the slower part (20–25 Hz) and the faster part (33–40 Hz) was plotted as a function of DP distance. Each dot refers to one tetrode. Red, distal one-third; green, intermediate; blue, proximal one-third. Low/high power ratio was negatively correlated with DP distance ( $r(38) = -0.50$ ,  $P < 0.001$ ). **e**, Time-resolved coherence spectrum for dCA1 versus pCA1. Data were averaged over 28 recording pairs in dCA1 and pCA1 from animals with tetrodes in both regions. **f**, Mean coherence between dCA1 and pCA1. Note lack of change in 20–40-Hz coherence during cue sampling compared to pre-cue or run periods (repeated measures ANOVA,  $F(2, 54) = 0.75$ ,  $P = 0.48$ ).





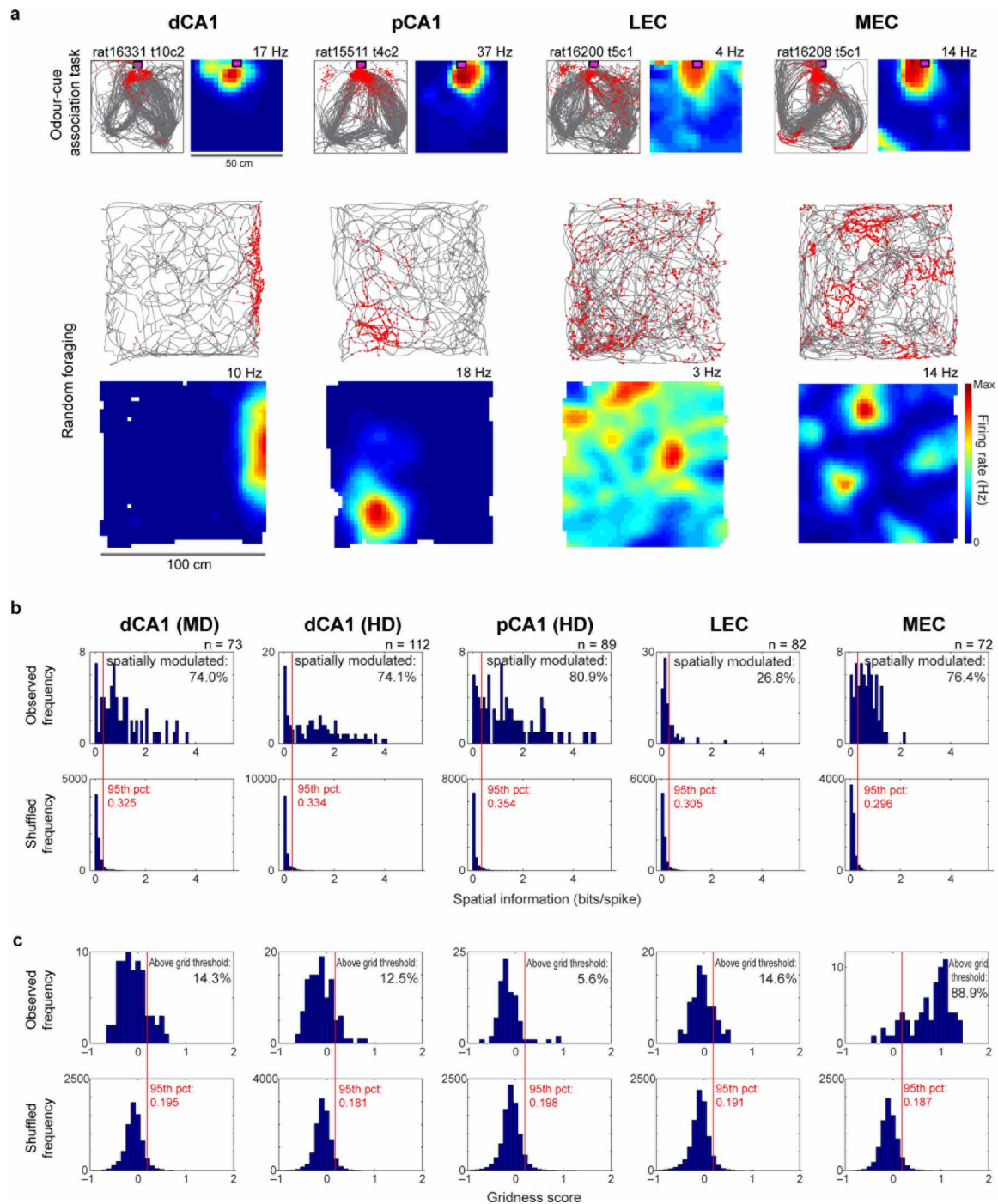
**Extended Data Figure 6 | Comparison of power and coherence development during learning.** **a, b**, Power of 20–40-Hz oscillations in dCA1 (**a**,  $n = 10$  tetrodes) and LEC (**b**,  $n = 10$  tetrodes) from 5 rats at time points T1–T5, on error trials at T5 (T5e) and at T5 after down-sampling (T5d), shown as values normalized by power at time point T1. Power of 20–40-Hz oscillations did not change significantly during the course of learning (repeated measures ANOVA:  $F(4, 36) = 0.72$ ,  $P = 0.58$  for dCA1;  $F(4, 36) = 1.71$ ,  $P = 0.17$  for LEC). Power on error trials (T5e) was not significantly different from correct trials (T5d; two-tailed paired  $t$ -test,  $t(9) = 1.40$ ,  $P = 0.20$  for dCA1;  $t(9) = 0.06$ ,  $P = 0.94$  for LEC). **c, d**, Power of theta oscillations in dCA1 (**c**) and LEC (**d**). Theta power did not change during learning (repeated measures ANOVA:  $F(4, 36) = 0.90$ ,  $P = 0.47$  for dCA1;  $F(4, 36) = 0.63$ ,  $P = 0.64$  for LEC). Theta power on error trials was not different from correct trials (two-tailed paired  $t$ -test,  $t(9) = 0.24$ ,  $P = 0.82$  for dCA1;  $t(9) = 0.40$ ,  $P = 0.70$  for LEC). **e**, Power of 20–40-Hz oscillations in dCA1 (left,  $n = 8$  tetrodes) and LEC (right,  $n = 8$  tetrodes) at the end of training with the original odours (A/B; time point T5), and with odours C/D and E/F (mean  $\pm$  s.e.m.) (4 rats). Power of the 20–40-Hz oscillations did

not change significantly with the new odours, neither between T5 and the first day with C/D or E/F ( $t(7) < 0.5$ ,  $P > 0.6$  for CA1;  $t(7) < 0.9$ ,  $P > 0.4$  for LEC) nor during the course of learning with C/D and E/F (repeated measures ANOVA:  $F(6, 42) = 0.34$ ,  $P = 0.91$  for dCA1;  $F(6, 42) = 0.44$ ,  $P = 0.85$  for LEC). **f**, Power of theta oscillations in dCA1 (left) and LEC (right) (mean  $\pm$  s.e.m.). There was no significant change in the power of theta (repeated measures ANOVA:  $F(6, 42) = 0.13$ ,  $P = 0.99$  for dCA1;  $F(6, 42) = 0.71$ ,  $P = 0.64$  for LEC). **g**, Mean of LEC–dCA1 coherence spectra during cue sampling at successive time points during learning (T1–T5) and on error trials (T5e) and down-sampled correct trials (T5d), plotted in the same way as the time-resolved coherence spectra in Fig. 2b ( $n = 20$  recording pairs from 5 rats). Dots above spectra denote frequencies with significant difference (FDR corrected for 62 multiple comparisons at 0–60 Hz,  $q < 0.05$ ). **h**, Coherence averaged across the theta-frequency band during cue sampling in novel odour trials. No change was observed for coherence in the theta frequency band (repeated measures ANOVA:  $F(2, 30) = 0.02$ ,  $P = 0.98$  for cues C/D and  $F(2, 30) = 0.25$ ,  $P = 0.78$  for cues E/F).



**Extended Data Figure 7 | Cross-frequency coupling of 20–40-Hz oscillations to local theta oscillations in LEC and pCA1.** **a–d**, Relationship of 20–40-Hz oscillations in LEC to phase of local theta oscillations, plotted as in Fig. 2d–g ( $n = 10$  tetrodes from 5 rats). **a**, Theta phase distribution of 20–40-Hz oscillation maxima at T1, T3, T5 and on error trials at T5 (T5e).  $0^\circ$  was defined as the trough of the theta cycle. Note that theta oscillations exist in LEC and that 20–40-Hz oscillations are moderately phase-coupled with theta oscillations already at T1. **b**, Mean vector length calculated from theta phase distribution of 20–40-Hz oscillation maxima at T1–T5. The degree of cross-frequency coupling did not change significantly across the learning period (T1–T5, repeated measures ANOVA:  $F(4, 36) = 1.7$ ,  $P = 0.17$ ). No difference was observed on error trials (T5e, compared with T5d using a two-tailed paired  $t$ -test,  $t(9) = 0.78$ ,  $P = 0.45$ ). **c**, Representative cross-frequency coherence plot showing for LEC that power of 20–40 Hz oscillations ( $y$ -axis) is modulated by theta phase ( $x$ -axis) during cue sampling at T5. Coupling strength is colour-coded (dark blue, no coupling; red, maximal coupling). **d**, Top, time-resolved power spectrum averaged across all theta cycles with 20–40-Hz oscillations at T5 in LFP from 10 tetrodes in all 5 rats.  $t = 0$  corresponds to the theta trough. Bottom, averaged unfiltered theta cycle. 20–40-Hz oscillations occurred at the falling phase of the theta wave. **e–h**, Similar plots for pCA1 ( $n = 14$  tetrodes

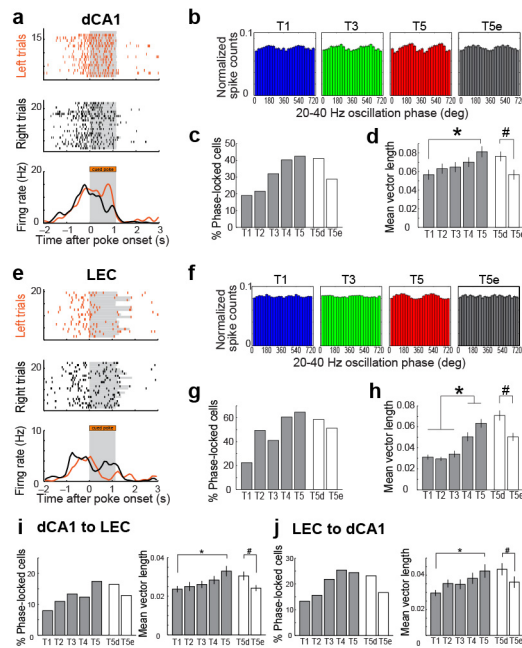
from 5 rats; **e**, theta phase distribution; **f**, mean vector length; **g**, cross-frequency coherence plot; **h**, time-resolved power spectrum). Animals with implants in pCA1 were recorded only after the completion of learning, that is, only at T5. In pCA1, oscillations at  $\sim 30$ –50-Hz were phase-coupled with theta oscillations. The degree of coupling did not change on error trials (two-tailed paired  $t$ -test,  $t(13) = 1.29$ ,  $P = 0.22$ ). **i**, Same plots as in Fig. 2f, but with wider  $60^\circ$  bins. The diagram shows the relationship of 20–40-Hz oscillations in dCA1 to the phase of local theta oscillations ( $n = 10$  tetrodes from 5 rats). The use of wider bins did not change the results. Significant cross-correlations between cell pairs in LEC and CA1 were not found, as expected due to the sparse connectivity between cell pairs in these areas<sup>13</sup>. **j**, Theta phase distribution of 20–40-Hz oscillation maxima for LFP from dCA1 as in Fig. 2f, but during the pre-cue period, at time points T1, T3, T5 ( $n = 10$  tetrodes from 5 rats, means  $\pm$  s.e.m.). **k**, Mean vector length calculated from theta phase distributions of 20–40-Hz oscillation maxima during the pre-cue period did not change during the course of training (repeated measures ANOVA:  $F(4, 36) = 0.26$ ,  $P = 0.90$ ;  $n = 10$  tetrodes from 5 rats). Vector lengths during cue sampling was increased compared to the pre-cue period at T5 (paired  $t$ -test,  $P < 0.05$ ;  $t(9) = 2.5$ ) but not at T1–T4 ( $P > 0.05$ ;  $t(9) < 1.9$ ).



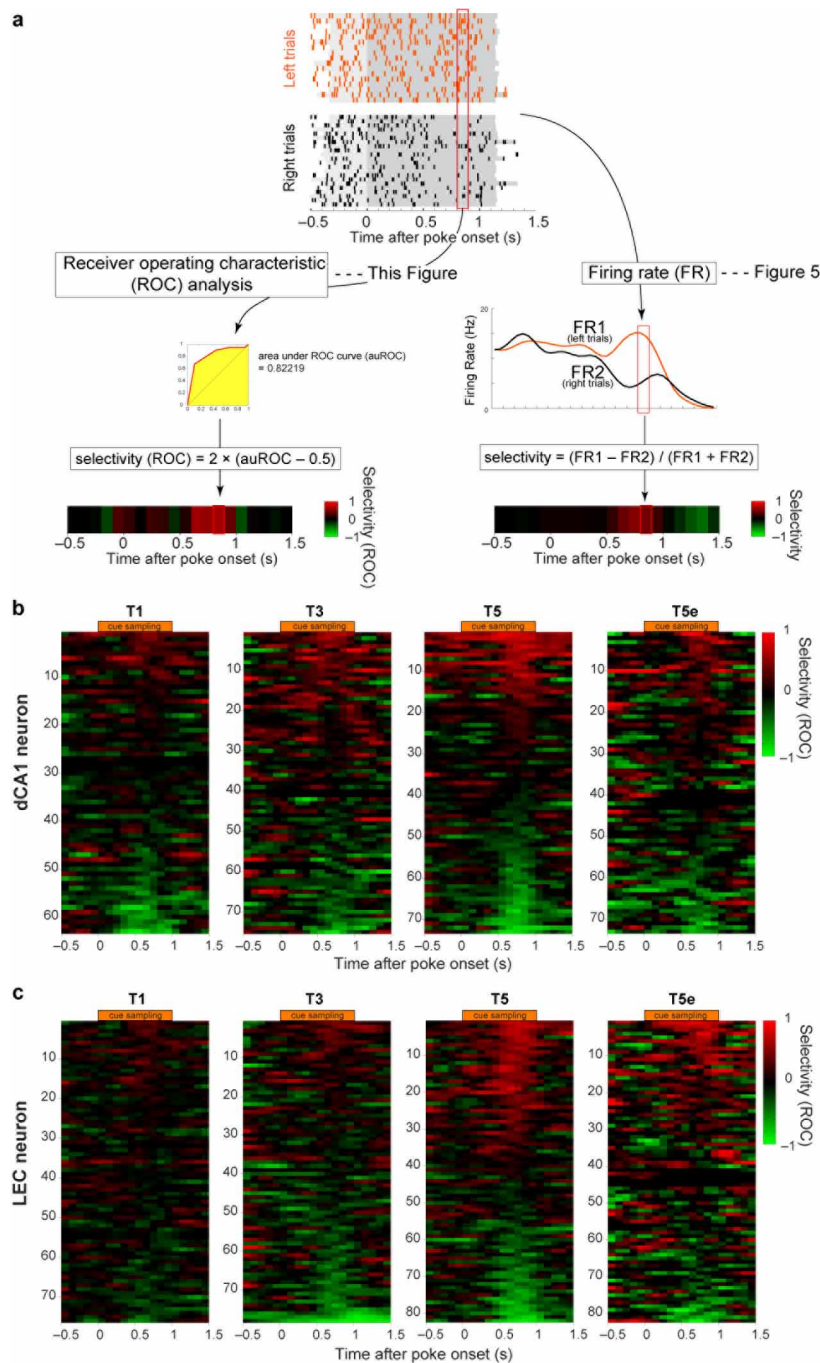
**Extended Data Figure 8 | A large fraction of CA1 and MEC cells with activity at the cue port were place cells and grid cells, respectively.** **a**, Spatial distribution of firing in example cells with cue-port activity recorded in dCA1, pCA1, LEC or MEC in either the odour-cue association task (top) or a random foraging task in a 1 m square box (bottom). Each column shows results for one representative cell. For each cell, spike position (red) is overlaid on the trajectory of the rat (grey) at the top and a colour-coded frequency map is shown at the bottom. Red is maximum firing rate, as indicated by the scale bar. Rat number, tetrode number (*t*) and cell number (*c*) are indicated above each path diagram. Peak frequencies are indicated at the top right of the colour map. Note that dCA1/pCA1 and MEC cells had clearly distinguishable place fields or grid fields in the foraging task, whereas the LEC cell showed no clear spatial

modulation. **b**, Top, distribution of spatial information scores calculated from firing rate distributions in the random foraging task for cells with cue-port activity in dCA1, pCA1, LEC and MEC. Results in dCA1 are shown for both microdrive (MD) and hyperdrive (HD) implants. Bottom, distribution of shuffled data based on 100 permutations per cell. Red lines indicate 95th percentile value (chance level) for a distribution based on all permutations in each area. 95th percentile value is indicated in red. Percentage of cells that exceeded chance level is shown for each region. Note that 74% and 81% of the cue-port cells were spatially modulated in dCA1 and pCA1, respectively, that is, they were place cells<sup>45</sup>. **c**, As in **b**, but for the distribution of gridness scores. Note that 89% of cue port cells in MEC had gridness scores that exceeded chance level and so were defined as grid cells<sup>45</sup>.





**Extended Data Figure 9 | Phase-locking of individual neurons to the 20–40-Hz rhythm.** **a**, Raster plots showing cue-port activity of example dCA1 cell on successive trials at T5. Rows correspond to individual trials; ticks indicate spikes. Top, trials with food in left cup; middle, trials with food in right cup; bottom, peri-stimulus time histogram (PSTH); orange, left-predicting trials; black, right-predicting trials. Between T1 and T5, the number of dCA1 cells with mean rates above 1 Hz during the cue interval ranged from 63 to 75 (5 rats). **b**, Spike-time distribution for dCA1 principal cells with cue-port activity across phase of local 20–40-Hz oscillation at T1–T5 and T5e. **c**, **d**, Percentage of significantly phase-locked cells (**c**) and mean vector length of distribution in **b** (**d**), averaged across dCA1 cells with cue-port activity. T5d as in Fig. 2c. \*#, as in Fig. 2c. **e–h**, LEC cells, as in **a–d**. Between T1 and T5, the number of LEC cells with mean rates above 1 Hz during the cue interval ranged from 76 to 82 (5 rats). The number of cue-port cells phase-locked to the local 20–40-Hz LFP increased significantly from T1 to T5 in both dCA1 and LEC (main text) and there was a significant increase in the phase locking of each cell (ANOVA for mean vector length: dCA1:  $F(4, 348) = 2.81$ ,  $P = 0.026$ ; LEC:  $F(4, 383) = 21.0$ ,  $P < 0.001$ ). On error trials, the mean vector for the spike-phase distribution decreased in both dCA1 and LEC (dCA1: two-tailed paired  $t$ -test,  $t(72) = 2.97$ ,  $P = 0.004$ ; LEC:  $t(81) = 4.05$ ,  $P < 0.001$ ). **i**, **j**, Interregional phase-locking of individual dCA1 and LEC cells. **i**, Phase-locking of dCA1 spikes as shown in **c** and **d**, but against 20–40-Hz oscillations in LEC. Left, percentage of significantly phase-locked cells. Right, mean vector length of the spike-phase distribution. The percentage of cells that was significantly phase-locked to the oscillations increased from 7.9 at T1 to 17.4 at T5 ( $P < 0.005$ ; binomial test with Bonferroni correction). This increase was matched by an increase in the mean vector length of the spike-phase distribution (ANOVA for mean vector length:  $F(4, 348) = 2.8$ ,  $P = 0.03$ ). T5e and T5d indicate T5 error trials and down-sampled correct T5 trials, respectively. Mean vector length on T5e was significantly reduced compared to T5d (two-tailed paired  $t$ -test,  $t(72) = 2.60$ ,  $P = 0.011$ ). **j**, Phase-locking of LEC spikes as shown in **g** and **h**, but against 20–40-Hz oscillations in dCA1. The percentage of LEC cells significantly phase-locked to 20–40-Hz oscillations in dCA1 increased from 13.3 at T1 to 24.4 at T5 ( $P < 0.005$ ). This was accompanied by a significant increase in the mean vector length of the spike-phase distribution ( $F(4, 383) = 2.7$ ,  $P = 0.03$ ). Mean vector length on error trials (T5e) decreased significantly compared to down-sampled correct trials (T5d; two-tailed paired  $t$ -test,  $t(81) = 2.42$ ,  $P = 0.017$ ). \* $P < 0.05$ , Bonferroni post-hoc test; # $P < 0.05$ , paired  $t$ -test.



**Extended Data Figure 10 | ROC-based and firing rate-based analysis of odour-specific representations in dCA1 and LEC.** **a**, Schematic representation of odour-type spike representations based on receiver operating characteristic (ROC) analysis (left) and direct comparison of firing rates (right). Odour-specific representation was assessed using the metric Selectivity. Selectivity was computed by comparing firing rates during sampling periods for left-associated and right-associated cues in successive 100-ms bins: Selectivity was first expressed as  $(\text{FR1} - \text{FR2}) / (\text{FR1} + \text{FR2})$ , where FR1 and FR2 are mean firing rates for multiple trials on individual 100-ms bins during left- and right-associated cue samplings, respectively. To confirm the

development of odour-specific representations, we subsequently computed Selectivity using ROC analysis, a method based on signal detection theory: Selectivity (ROC) was computed as  $2 \times (\text{auROC} - 0.5)$ , where auROC is the area under the ROC curve computed from spike numbers for left and right trials on individual 100-ms bins. After scaling, for both Selectivity metrics, Selectivity = 1 indicates that the cell fired only on left-associated odour trials, whereas Selectivity = -1 denotes firing only on right-associated odour trials. **b**, Trial-type representations for all dCA1 cells with activity at the cue port shown as in Fig. 4a but using ROC analysis. Note development of selectivity also with this method. **c**, As in **b**, but for LEC cells.

# Trace-gas metabolic versatility of the facultative methanotroph *Methylocella silvestris*

Andrew T. Crombie<sup>1</sup> & J. Colin Murrell<sup>1</sup>

The climate-active gas methane is generated both by biological processes and by thermogenic decomposition of fossil organic material, which forms methane and short-chain alkanes, principally ethane, propane and butane<sup>1,2</sup>. In addition to natural sources, environments are exposed to anthropogenic inputs of all these gases from oil and gas extraction and distribution. The gases provide carbon and/or energy for a diverse range of microorganisms that can metabolize them in both anoxic<sup>3</sup> and oxic zones. Aerobic methanotrophs, which can assimilate methane, have been considered to be entirely distinct from utilizers of short-chain alkanes, and studies of environments exposed to mixtures of methane and multi-carbon alkanes have assumed that disparate groups of microorganisms are responsible for the metabolism of these gases. Here we describe the mechanism by which a single bacterial strain, *Methylocella silvestris*, can use methane or propane as a carbon and energy source, documenting a methanotroph that can utilize a short-chain alkane as an alternative to methane. Furthermore, during growth on a mixture of these gases, efficient consumption of both gases occurred at the same time. Two soluble di-iron centre monooxygenase (SDIMO) gene clusters were identified and were found to be differentially expressed during bacterial growth on these gases, although both were required for efficient propane utilization. This report of a methanotroph expressing an additional SDIMO that seems to be uniquely involved in short-chain alkane metabolism suggests that such metabolic flexibility may be important in many environments where methane and short-chain alkanes co-occur.

Most of the 500–600 Tg methane emitted into the atmosphere each year is of recent biological origin, whereas fossil-derived ‘natural gas’, from both anthropogenic and natural sources, contributes approximately 30% of the total<sup>4</sup>. Large anthropogenic releases of natural gas have included the Deepwater Horizon spill of 2010 (ref. 5), but intentional or unintentional operational releases are also widespread and likely to increase with the exploitation of unconventional resources, including shale gas extraction (fracking)<sup>6</sup>. A spectacular natural release, containing methane and 35% (v/v) ethane and propane<sup>7</sup>, results in the ‘eternal flame’ in Chestnut Ridge Park, New York. Widespread seepage, which is largely undocumented, also occurs in terrestrial areas that overlay sedimentary organic carbon<sup>8</sup>. The natural geological methane source, including terrestrial macro-seeps and micro-seeps, marine seeps and volcanic and geothermal emissions, is the second largest natural source after wetlands. Micro-seeps, which are potentially active in substantial areas (including petroliferous regions and sedimentary basins where thermal degradation of organic material has occurred), contribute 10–25 Tg methane annually<sup>8</sup>. This thermogenic natural gas contains methane and up to 50% (w/w) ethane, propane and butane<sup>1,2</sup>. In addition, unquantified biogenic production of ethane and propane occurs in anoxic environments similar to those that support methanogenesis and homoacetogenesis<sup>9</sup>.

Aerobic methanotrophs oxidize much of the methane that is released from terrestrial and marine environments before it reaches the atmosphere and are well characterized<sup>10</sup>, and microorganisms that grow on short-chain alkanes, including propane, have also received attention<sup>11–15</sup>.

However, so far, all evidence has emphasized the striking distinction between the groups of microorganisms performing these respective tasks. Aerobic methanotrophs are Gram-negative bacteria from about 16 genera within the Gammaproteobacteria and Alphaproteobacteria classes, as well as some extremophiles of the phylum Verrucomicrobia. Two types of enzyme that can oxidize methane are known: the membrane-associated particulate methane monooxygenase (pMMO), which was previously thought to be present in all methanotrophs; and a cytoplasmic, soluble form of this enzyme (sMMO), which is also present in a few extant strains. The sMMOs form one group of a large family of di-iron carboxylate enzymes that includes SDIMOs that allow growth on a range of short-chain alkanes, alkenes and aromatic compounds. Recently, a few methanotrophs were shown to have restricted growth on two-carbon compounds, including acetate<sup>16</sup>; however, there are no validated reports of growth on (as opposed to co-oxidation of) gaseous hydrocarbons other than methane. Short-chain alkane utilizers have been thought to be incapable of methane metabolism, and most are members of the class Actinobacteria<sup>12</sup>. An exception is *Thauera butanivorans*, which is a betaproteobacterium that grows on butane using a monooxygenase that is two to three orders of magnitude more specific for butane than for methane<sup>17</sup>, despite its sequence similarity (up to 65%) to bona fide methane monooxygenases.

Methanotrophic strains of the genus *Methylocella*, first isolated from peat, tundra and forest soils in Northern Europe<sup>18</sup>, are widespread in a diverse range of terrestrial environments<sup>19</sup> and have also been found, for example, among the microbial community following the Deepwater Horizon incident<sup>20</sup>. The ability of *M. silvestris* to grow on organic acids and alcohols is known, as are details of its sMMO<sup>18,21</sup>. Here we further investigated the metabolic potential of this organism (Extended Data Table 1). *M. silvestris* grew on propane (2.5% or 20% (v/v)) at a similar rate to its growth solely on methane in batch culture (Table 1), reaching a high cell density in a fermenter supplied with 20% (v/v) propane in air (Extended Data Fig. 1). It also grew on potential intermediates of propane oxidation: propionate, 2-propanol and acetone. Its growth on methane was not inhibited by the presence of 1% or 10% (v/v) propane in the methane, and indeed the conversion of substrate carbon into biomass was slightly enhanced under both of these conditions of dual substrate

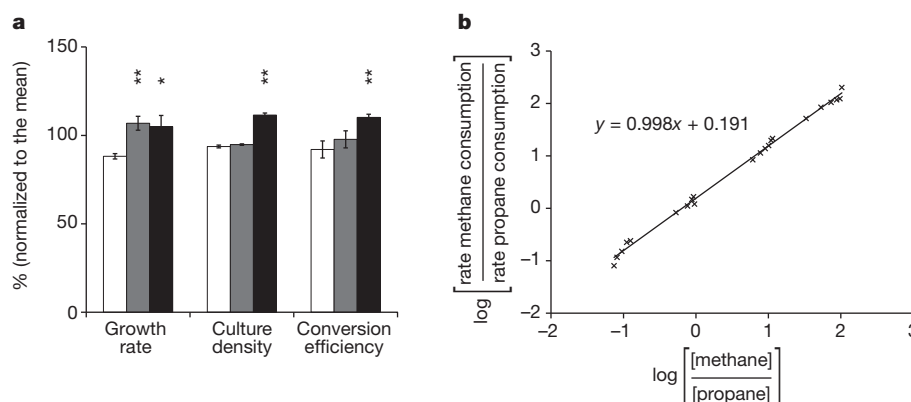
**Table 1 | Growth rates of *Methylocella silvestris* strains**

Strain	Substrate	Growth rate (h <sup>-1</sup> )	
		On 2.5% (v/v)	On 20% (v/v)
Wild-type	Methane	0.012 ± 0.001	0.013 ± 0.003
	Propane	0.005 ± 0.000	0.015 ± 0.002
	Mixture	0.011 ± 0.001	ND
ΔMmoX	Methane	No growth	No growth
	Propane	No growth	0.010 ± 0.000
	Mixture	No growth	ND
ΔPrmA	Methane	0.010 ± 0.003	0.019 ± 0.005
	Propane	0.003 ± 0.000	No growth
	Mixture	0.010 ± 0.000	ND

Specific growth rates of wild-type *M. silvestris* and the ΔMmoX and ΔPrmA strains during growth on methane or propane (at the concentrations shown) or a mixture (2.5% (v/v) each). The data are the mean ± s.d. of triplicate vials (except for 2.5% (v/v) propane, for which duplicate vials were used). ND, not determined.

<sup>1</sup>School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK.





**Figure 1 | Propane enhanced *Methylocella silvestris* growth, and methane and propane were consumed by cultures at rates corresponding to their relative concentrations.** **a**, The specific growth rate ( $\mu$ ), maximum culture density (optical density at 540 nm ( $OD_{540}$ )) and carbon conversion efficiency (CCE) of cultures supplied with equimolar amounts of carbon as methane (10% (v/v), white) or mixtures of methane and propane (99:1 (v/v), grey; 90:10

(v/v), black). The data are normalized to the mean ( $\mu = 0.018 \text{ h}^{-1}$ ,  $OD_{540} = 0.36$  and  $CCE = 0.28 \text{ mg protein per mg substrate carbon}$ );  $n = 3$  vials per growth condition. Error bars, s.d. Significance was determined by single factor analysis of variance (ANOVA) (compared with growth on 100% methane): \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ . **b**, Relative rates of methane and propane consumption by cultures supplied with a range of gas mixtures.

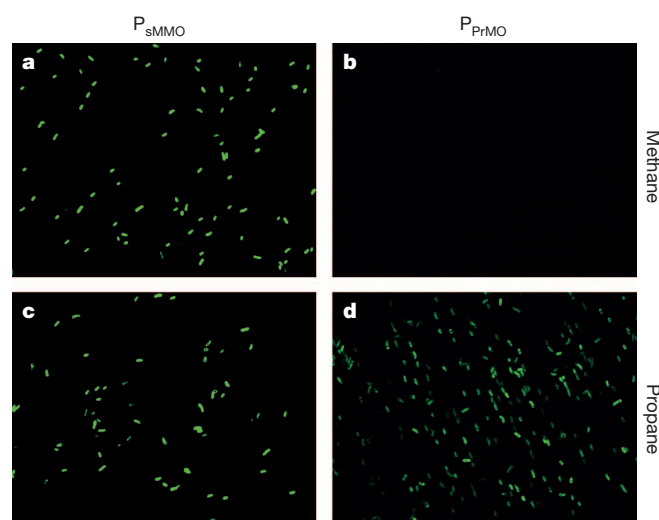
availability (Fig. 1a). To identify any substrate preference, liquid cultures of *M. silvestris* supplied with various mixtures of methane and propane were monitored. The components of the gas mixtures, ranging from 1:10 to 100:1 (methane:propane), were differentially consumed in direct proportion to their relative headspace concentrations (Fig. 1b), with methane being consumed at a volumetric rate approximately 1.5 times that of propane when supplied at equivalent concentrations. The similar Henry's law coefficients<sup>22</sup> of methane and propane ( $K_H = 1.3 \times 10^{-3}$  and  $1.4 \times 10^{-3} \text{ M atm}^{-1}$ , respectively) and consequent aqueous concentrations indicated that the overall kinetic parameters in these microcosms were similar and that under these conditions *M. silvestris* did not exhibit a major substrate preference for either of these gases.

Analysis of the genome of *M. silvestris*<sup>23</sup> confirmed the absence of the membrane-associated pMMO but revealed an SDIMO gene cluster in addition to that encoding sMMO. Phylogenetic analysis of the putative hydroxylase  $\alpha$ -subunit grouped this second SDIMO with those of propane-oxidizing bacteria, including *Gordonia* sp. TY-5 (ref. 14). These SDIMOs belong to Group V SDIMO enzymes as defined by Coleman and co-workers<sup>24</sup> (Extended Data Fig. 2a), suggesting that in *M. silvestris* this gene cluster might encode a propane monooxygenase (PrMO).

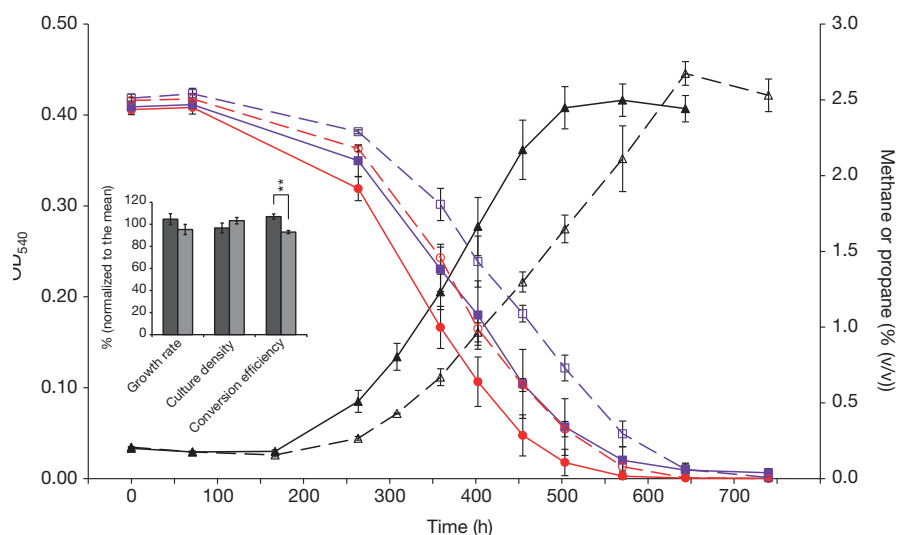
Transcription of this SDIMO was confirmed by reverse-transcription PCR (RT-PCR); 5' rapid amplification of cDNA ends (RACE) was used to identify the transcription start site, which is 120 nucleotides upstream (5') of the predicted start codon, and a putative  $\sigma^{54}$  promoter was located (Extended Data Fig. 2b). A reporter gene (green fluorescent protein (GFP)) under the control of each of the putative promoters of these two enzyme systems (sMMO and PrMO) was then expressed in *M. silvestris* from plasmids: the sMMO genes, but not the PrMO genes, were transcribed during growth on methane, but both sets of genes were transcribed during growth on propane (Fig. 2). The protein profiles differed between these two growth conditions, and mass spectrometric analysis of bands excised from a gel confirmed that protein expression mirrored gene transcription (Extended Data Fig. 3a, b).

The sMMO from other methanotrophs can co-oxidize numerous substrates, including propane<sup>25</sup>, although these organisms cannot assimilate the oxidation products, whereas the specificity of PrMO is relatively unknown. Therefore, although methanotrophs have been notoriously difficult to genetically manipulate, two deletion mutants were constructed by targeted marker-exchange mutagenesis, to unravel the relative contribution of each enzyme to growth on alkanes. Wild-type *M. silvestris* and *M. silvestris* strain  $\Delta\text{MmoX}$  and strain  $\Delta\text{PrmA}$  (with deletions of the hydroxylase  $\alpha$ -subunits of sMMO and PrMO, respectively) were grown in batch culture on 20% (v/v) methane or 20% (v/v) propane. At this substrate concentration, the deletion strains did not grow on the gas for

which they lacked the nominal oxidizing enzyme (Table 1). When cultures were supplied with 2.5% (v/v) methane or 2.5% (v/v) propane, the wild-type strain grew on both substrates. By contrast, the  $\Delta\text{MmoX}$  strain did not grow on 2.5% (v/v) methane, 2.5% (v/v) propane or a mixture of these (2.5% (v/v) of each). The  $\Delta\text{PrmA}$  strain grew on 2.5% (v/v) methane similarly to the wild-type and, interestingly, also on 2.5% (v/v) propane but at a reduced rate (Table 1) and with a 30% reduction in the conversion of the substrate carbon into biomass compared with the wild-type strain (Extended Data Fig. 4a). During growth on a mixture of methane and propane (2.5% (v/v) each), only a marginal difference in growth rate was observed between the wild-type and  $\Delta\text{PrmA}$  strains, and the  $\Delta\text{PrmA}$  strain consumed both gases simultaneously, similarly to the wild-type strain, although a small decrease in the carbon conversion efficiency was evident (Fig. 3). We also noted that the carbon source (succinate or methane plus propane) for the cells used as the inoculum did not affect a culture's subsequent simultaneous oxidation of methane and propane (Extended Data Fig. 4b).



**Figure 2 | Methane induced transcription of sMMO only, whereas propane induced both sMMO and PrMO.** Fluorescence microscopy images of *M. silvestris* cells grown to late exponential phase on methane alone (a, b) or propane alone (c, d) transformed with plasmids containing the sMMO promoter (a, c) or the PrMO promoter (b, d) fused to a GFP (green) reporter gene. The images are representative of one of two independent experiments. Cells are approximately  $1.5 \mu\text{m}$  in length.



**Figure 3 | Inactivation of PrMO had a marginal effect during growth on a mixture of methane and propane.** Growth (left y axis, black lines) and substrate consumption (right y axis, coloured lines) of wild-type *M. silvestris* (solid lines, filled symbols) and the  $\Delta$ PrmA strain (dashed lines, open symbols). The inoculum was grown on succinate. The experimental cultures were supplied with a mixture of methane and propane (2.5% (v/v) each). The headspace concentrations of methane (red) and propane (purple) were determined by gas chromatography.  $n = 3$  vials per strain. Error bars, s.d. The inset shows the growth parameters for the wild-type strain (dark grey) and the  $\Delta$ PrmA strain (light grey). The inset data are normalized to the mean ( $\mu = 0.010 \text{ h}^{-1}$ ,  $\text{OD}_{540} = 0.43$  and  $\text{CCE} = 0.30 \text{ mg protein per mg substrate carbon}$ );  $n = 3$  vials per strain. Error bars, s.d. Significance was determined by Student's *t*-test: \*\*,  $P < 0.01$ .

Because sMMO can oxidize both methane and propane, the converse ability of PrMO, to oxidize methane, was evaluated by comparing methane consumption by the wild-type and  $\Delta$ MmoX strains during growth on 20% (v/v) propane in the presence of 2% (v/v) methane. The wild-type strain consumed both gases in proportion to their relative concentrations, as expected, whereas there was no detectable consumption of methane by the  $\Delta$ MmoX strain (Extended Data Fig. 4c, d).

The relative inability of PrMO to oxidize methane was used to interpret the comparative substrate oxidation rates by whole cells of wild-type *M. silvestris*, as obtained by polarographic measurement of oxygen uptake in response to the addition of methane or propane, in an oxygen electrode. The maximum methane oxidation rates of cells grown on methane (expressing only sMMO) were 2.3-fold higher than those of cells grown on propane. By contrast, the maximum propane oxidation rates were 2.4-fold higher in propane-grown cells (expressing both sMMO and PrMO) than in methane-grown cells, indicating that in propane-grown cells, the propane-oxidizing ability was due mainly to PrMO (Extended Data Fig. 5a). Using this technique, methane-grown cells had an apparent Michaelis constant ( $K_m$ ) of 53  $\mu\text{M}$  for methane, and propane-grown cells had an apparent  $K_m$  of 19  $\mu\text{M}$  for propane (Extended Data Fig. 5b, c). The relatively high affinity of PrMO for propane emphasizes its important role in propane oxidation and is consistent with the reduced growth rate of the  $\Delta$ PrmA strain on this substrate. Therefore, *M. silvestris* contains two enzymes capable of propane oxidation, sMMO and PrMO, and deletion of PrMO is not sufficient to completely eliminate the oxidation of propane.

Bacterially mediated propane oxidation typically occurs at the terminal or sub-terminal carbon, producing 1-propanol or 2-propanol, respectively, or a mixture<sup>12</sup>. Terminal oxidation proceeds from 1-propanol to propanal and propionate, whereas 2-propanol is oxidized to acetone. Oxygen electrode assays demonstrated that, in comparison with cells grown on methane or succinate, propane-grown cells consumed threefold or fivefold more oxygen, respectively, in response to addition of 1-propanol and propanal (Extended Data Fig. 6a), and growth tests indicated that 1-propanol-metabolizing ability was induced in propane-grown cells (Extended Data Fig. 7a–d). Quantitative label-free mass-spectrometric proteomic analyses showed that polypeptides of the methylmalonyl-CoA pathway of propionate oxidation were induced during growth on propane<sup>26</sup>, indicating terminal oxidation. However, 2-propanol was detected in culture media, reaching a maximum of approximately 0.5 mM at 200 h during growth on 4% (v/v) propane, demonstrating that at least 25% of propane was oxidized at the sub-terminal position (Extended Data Fig. 8). When propane was depleted, 2-propanol was consumed by the cultures, and enzyme activities associated with the oxidation of sub-terminal intermediates were upregulated in propane-grown cells (Extended Data Fig. 6b, c). Therefore, in *M. silvestris*, propane was oxidized

to both 1-propanol and 2-propanol. However, whereas 2-propanol (reaching 1.3 mM at 1,050 h) was also detected in the medium of the  $\Delta$ PrmA strain (grown on 4% (v/v) propane) (Extended Data Fig. 9a), it was not detected in the medium of the  $\Delta$ MmoX strain (grown on 20% (v/v) propane). The sMMO of obligate methanotrophs has previously been shown to oxidize *n*-alkanes to a mixture of primary and secondary alcohols<sup>25</sup>, and our data suggest that in *M. silvestris*, 2-propanol is a product of sMMO but not a major product of PrMO.

We therefore examined the protein expression pattern of the wild-type strain during growth on 2-propanol (Extended Data Fig. 3), and we observed that, under these conditions, PrMO, but not sMMO, was expressed. Therefore, these data suggest that propane oxidation by sMMO is sufficient for the induction of PrMO. The inability of the  $\Delta$ MmoX strain to grow at low propane concentrations probably resulted from a lack of sub-terminal metabolites, which may be responsible for the induction of PrMO expression. Interestingly, deletion of *prmA* also halved the growth rate of cultures on 2-propanol but had no significant effect on growth on methanol, ethanol, acetate or acetone (Extended Data Fig. 9b). A similar unexplained phenomenon has been noted by others<sup>14,27</sup> after the genetic inactivation of propane-oxidizing enzymes in *Gordonia* and *Mycobacterium* spp. The decreased oxidizing ability, and consequent increased accumulation, of 2-propanol in the  $\Delta$ PrmA strain would also have contributed to its growth rate and carbon conversion efficiency being lower than those of the wild-type during growth on low concentrations of propane. Similarly, during incubation of this strain with higher propane concentrations, the 2-propanol that formed as a result of propane oxidation by sMMO may have accumulated to toxic levels, thereby inhibiting growth. This is consistent with PrMO having a role in the oxidation of 2-propanol, either directly or through a regulatory mechanism. To identify the ability of PrMO to directly participate in 2-propanol oxidation, and because acetylene is a potent irreversible inhibitor of SDIMOs<sup>28</sup>, we grew the wild-type strain on propane, 2-propanol or acetone in the presence or absence of 2% (v/v) acetylene. Growth on propane was entirely prevented by the presence of acetylene, whereas growth on acetone was unaffected. Although this inhibition of PrMO reduced the specific growth rate of cultures supplied with 2-propanol by approximately 11% compared with uninhibited controls (Extended Data Fig. 9c), this reduction was far less than that caused by the gene deletion in the  $\Delta$ PrmA strain, suggesting that the oxidation of 2-propanol by PrMO was relatively minor. Therefore, the exact mechanism by which PrMO is required for efficient 2-propanol metabolism remains to be elucidated.

To better understand the significance and role of facultative methanotrophs in the environment, we investigated the behaviour of whole cells in culture, rather than taking the more reductionist approach of analysing the specificities and kinetics of purified enzymes *in vitro*. The data indicated that sMMO was necessary and sufficient for growth on

methane but that both SDIMOs were required for efficient metabolism of propane. Although sMMO was capable of propane oxidation, it did not exclusively perform this role, and cells lacking PrMO were at a disadvantage that increased with increasing substrate concentration. This report is to our knowledge the first to describe an organism that can grow on methane or a short-chain alkane, disproving the dogma that separate microorganisms are responsible for the metabolism of methane and other simple hydrocarbon gases, such as propane. The *pmoA* gene, which encodes the  $\alpha$ -subunit of pMMO, has frequently been used as an environmental probe for methanotrophs<sup>29</sup>, raising the possibility that sMMO-only methanotrophs have been overlooked, and some of these might have similar capabilities to those described here. Given the widespread co-occurrence of methane and short-chain alkanes, including propane, in the environment, the potential for simultaneous oxidation of these gases may be considerable. This finding also has implications for studies of carbon cycling in environments where natural gas is released, such as the 8 million square kilometres of Earth's terrestrial regions, where micro-seeps may be abundant<sup>6</sup>, or near marine methane seeps, oil spills and gas extraction and distribution systems.

## METHODS SUMMARY

*Methylocella silvestris* BL2 was grown in 20 ml or 25 ml volumes in 120-ml serum vials sealed with rubber stoppers as previously described<sup>30</sup> and supplied with methane or propane by injection through the septum into the headspace. Headspace gas concentrations were quantified by gas chromatography, as were metabolites after solvent extraction from culture medium. The carbon conversion efficiency was calculated as the ratio of culture protein production to substrate carbon depletion. DNA and RNA procedures followed standard methods, and transcriptional start sites were determined using RACE. GFP-expressing reporter strains were constructed as previously described<sup>21</sup>. For gel-based protein analysis, bands of interest were excised, and polypeptides were identified by mass spectrometry. Mutant strains were constructed as previously described<sup>30</sup>. Polarographic activity assays used a Clark-type oxygen electrode to track substrate-induced oxygen consumption.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 13 June 2013; accepted 26 February 2014.**

**Published online 28 April; corrected online 1 May 2014 (see full-text HTML version for details).**

1. Etiope, G. & Ciccioli, P. Earth's degassing: a missing ethane and propane source. *Science* **323**, 478 (2009).
2. Mango, F. D., Hightower, J. W. & James, A. T. Role of transition-metal catalysis in the formation of natural gas. *Nature* **368**, 536–538 (1994).
3. Heider, J., Spormann, A. M., Beller, H. R. & Widdel, F. Anaerobic bacterial metabolism of hydrocarbons. *FEMS Microbiol. Rev.* **22**, 459–473 (1998).
4. Etiope, G., Lassey, K. R., Klusman, R. W. & Boschi, E. Reappraisal of the fossil methane budget and related emission from geologic sources. *Geophys. Res. Lett.* **35**, L09307 (2008).
5. Reddy, C. M. *et al.* Composition and fate of gas and oil released to the water column during the Deepwater Horizon oil spill. *Proc. Natl Acad. Sci. USA* **109**, 20229–20234 (2012).
6. Osborn, S. G., Vengosh, A., Warner, N. R. & Jackson, R. B. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proc. Natl Acad. Sci. USA* **108**, 8172–8176 (2011).
7. Etiope, G., Drobnik, A. & Schimmelmann, A. Natural seepage of shale gas and the origin of 'eternal flames' in the Northern Appalachian Basin, USA. *Mar. Petrol. Geol.* **43**, 178–186 (2013).
8. Etiope, G. & Klusman, R. W. Microseepage in drylands: flux and implications in the global atmospheric source/sink budget of methane. *Glob. Planet. Change* **72**, 265–274 (2010).
9. Hinrichs, K. U. *et al.* Biological formation of ethane and propane in the deep marine subsurface. *Proc. Natl Acad. Sci. USA* **103**, 14684–14689 (2006).
10. Reece, W. S. in *Treatise on Geochemistry* (eds Holland, H. D. & Turekian, K. K.) 1–32 (Elsevier, 2007).
11. Hamamura, N. & Arp, D. J. Isolation and characterization of alkane-utilizing *Nocardioides* sp. strain CF8. *FEMS Microbiol. Lett.* **186**, 21–26 (2000).
12. Ashraf, W., Mihdhir, A. & Murrell, J. C. Bacterial oxidation of propane. *FEMS Microbiol. Lett.* **122**, 1–6 (1994).
13. Kotani, T., Kawashima, Y., Yurimoto, H., Kato, N. & Sakai, Y. Gene structure and regulation of alkane monooxygenases in propane-utilizing *Mycobacterium* sp. TY-6 and *Pseudonocardia* sp. TY-7. *J. Biosci. Bioeng.* **102**, 184–192 (2006).
14. Kotani, T., Yamamoto, T., Yurimoto, H., Sakai, Y. & Kato, N. Propane monooxygenase and NAD<sup>+</sup>-dependent secondary alcohol dehydrogenase in propane metabolism by *Gordonia* sp. strain TY-5. *J. Bacteriol.* **185**, 7120–7128 (2003).
15. Johnson, E. L. & Hyman, M. R. Propane and *n*-butane oxidation by *Pseudomonas putida* GPo1. *Appl. Environ. Microbiol.* **72**, 950–952 (2006).
16. Semrau, J. D., DiSpirito, A. A. & Vuilleumier, S. Facultative methanotrophy: false leads, true results, and suggestions for future research. *FEMS Microbiol. Lett.* **323**, 1–12 (2011).
17. Cooley, R. B., Dubbels, B. L., Sayavedra-Soto, L. A., Bottomley, P. J. & Arp, D. J. Kinetic characterization of the soluble butane monooxygenase from *Thauera butanivorans*, formerly '*Pseudomonas butanovora*'. *Microbiology* **155**, 2086–2096 (2009).
18. Dedysh, S. N., Knief, C. & Dunfield, P. F. *Methylocella* species are facultatively methanotrophic. *J. Bacteriol.* **187**, 4665–4670 (2005).
19. Rahman, M. T. *et al.* Environmental distribution and abundance of the facultative methanotroph *Methylocella*. *ISME J.* **5**, 1061–1066 (2011).
20. Mason, O. U. *et al.* Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J.* **6**, 1715–1727 (2012).
21. Theisen, A. R. *et al.* Regulation of methane oxidation in the facultative methanotroph *Methylocella silvestris* BL2. *Mol. Microbiol.* **58**, 682–692 (2005).
22. Mackay, D. & Shiu, W. Y. A critical review of Henry's law constants for chemicals of environmental interest. *J. Phys. Chem. Ref. Data* **10**, 1175–1199 (1981).
23. Chen, Y. *et al.* Complete genome sequence of the aerobic facultative methanotroph *Methylocella silvestris* BL2. *J. Bacteriol.* **192**, 3840–3841 (2010).
24. Coleman, N. V., Bui, N. B. & Holmes, A. J. Soluble di-iron monooxygenase gene diversity in soils, sediments and ethene enrichments. *Environ. Microbiol.* **8**, 1228–1239 (2006).
25. Colby, J., Stirling, D. I. & Dalton, H. The soluble methane mono-oxygenase of *Methylococcus capsulatus* (Bath). Its ability to oxygenate *n*-alkanes, *n*-alkenes, ethers, and alicyclic, aromatic and heterocyclic compounds. *Biochem. J.* **165**, 395–402 (1977).
26. Patel, N. A. *et al.* Comparison of one- and two-dimensional liquid chromatography approaches in the label-free quantitative analysis of *Methylocella silvestris*. *J. Proteome Res.* **11**, 4755–4763 (2012).
27. Furuya, T., Hirose, S., Osanai, H., Semba, H. & Kino, K. Identification of the monooxygenase gene clusters responsible for the regioselective oxidation of phenol to hydroquinone in mycobacteria. *Appl. Environ. Microbiol.* **77**, 1214–1220 (2011).
28. Prior, S. D. & Dalton, H. Acetylene as a suicide substrate and active site probe for methane monooxygenase from *Methylococcus capsulatus* (Bath). *FEMS Microbiol. Lett.* **29**, 105–109 (1985).
29. McDonald, I. R., Bodrossy, L., Chen, Y. & Murrell, J. C. Molecular ecology techniques for the study of aerobic methanotrophs. *Appl. Environ. Microbiol.* **74**, 1305–1315 (2008).
30. Crombie, A. & Murrell, J. C. Development of a system for genetic manipulation of the facultative methanotroph *Methylocella silvestris* BL2. *Methods Enzymol.* **495**, 119–133 (2011).

**Acknowledgements** This work was funded by a Natural Environmental Research Council (NERC) grant (NE/E016855/1) and a NERC studentship to A.T.C. Additional funding was provided by the University of East Anglia and the Earth and Life Systems Alliance of the Norwich Research Park. We thank A. Johnston for critical reading of the manuscript.

**Author Contributions** A.T.C. designed and performed experiments, analysed data and wrote the paper. J.C.M. designed experiments and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.C.M. ([j.c.murrell@uea.ac.uk](mailto:j.c.murrell@uea.ac.uk)).



## METHODS

**Cultivation of *Methylocella silvestris*.** *Methylocella silvestris* BL2 was grown in dilute nitrate mineral salts (DNMS) medium (20 ml or 25 ml) in 120-ml serum vials sealed with rubber stoppers as previously described<sup>30</sup> and supplied with methane or propane by injection through the septum into the headspace. Methane and propane were a minimum of 99.5% purity grade. Large-scale cultivation was carried out in 4 l fermenters (LH Series 210, LH Engineering), or 2 l fermenters (FerMac 300, Electrolab) supplied with methane or propane (50–200 ml min<sup>-1</sup>) and air. Other substrates were used at a concentration of 5 mM (succinate and acetate) or 0.05% (v/v) (methanol, ethanol, 2-propanol and acetone) except where indicated. To calculate growth rates, the natural logarithm of culture density (as measured by optical density at 540 nm (OD<sub>540</sub>)) was plotted against time, and a straight line was fitted to the exponential phase of growth using Microsoft Excel. The growth rate was determined for each culture from a minimum of three data points or two data points separated by at least 48 h. The increase in biomass was defined as the difference between the initial and maximum densities (OD<sub>540</sub>) reached by the culture. The consumption of methane or propane was calculated from the depletion of headspace substrate between consecutive time points during the growth phase and was plotted against the mean of the concentrations at these points. The carbon conversion efficiency was evaluated as the ratio of protein production to substrate carbon depletion.

**Measurement of substrate depletion and product formation.** The headspace gas concentrations were quantified using an Agilent 6890A or 7820A gas chromatograph equipped with a Porapak Q column (Supelco) coupled to a flame ionization detector (FID), with injector, column and detector temperatures of 150 °C, 125 °C and 200 °C, respectively. Standard curves for methane and propane were prepared using standards containing known concentrations of the pure gases in air. The products of propane oxidation in the culture medium were quantified by solvent extraction. Aliquots (1.5 ml) of the culture were centrifuged (16,600 g, 5 min, room temperature (21 °C) to pellet the cells. The supernatant (800 µl) was added to 400 µl ethyl acetate (containing 500 parts per million volume (p.p.m.v.) 1-butanol as an internal standard) and approximately 0.4 g NaCl, (sufficient to saturate the aqueous phase), in 1.5 ml tubes. The samples were mixed for 30 min on a tube rotator and briefly centrifuged to separate the phases. An autosampler was used to inject 5 µl organic phase (100:1 split) into an Agilent 7890A gas chromatograph equipped with an HP-INNOWax column (30 m × 0.32 mm × 0.25 µm) and FID. The injector temperature was 250 °C; the detector was at 250 °C; and the oven was programmed to hold at 45 °C for 5 min, then to increase from 45 °C to 250 °C at 10 °C min<sup>-1</sup>, and then to hold at 250 °C for 5 min. The retention time for acetone was 2.40 min; 1-propanol, 3.34 min; 2-propanol, 5.40 min; 1-butanol, 7.45 min; and acetol, 9.70 min. Compounds of interest were quantified using aqueous standards extracted using the same method.

**Measurement of cellular protein.** Total cellular protein was measured following a previously reported method<sup>31</sup> with minor modifications. Aliquots of cultures (500 µl) were stored frozen for analysis. Following thawing, 87.5 µl 20% (w/v) SDS and 87.5 µl 1.6 M NaOH were added, and the tubes were heated in a boiling water bath for 5 min. The tubes were cooled, and the NaOH was neutralized by the addition of 87.5 µl 1.6 M HCl, centrifuged (13,000 g, 5 min, 21 °C), and 650 µl supernatant was removed to fresh tubes. The protein was assayed using a Pierce BCA kit (Fisher Scientific), following the manufacturer's recommendations and using standards (500 µl) containing 0–120 µg bovine serum albumin ml<sup>-1</sup>, treated in an identical manner to the samples.

**Phylogenetic analysis of *MmoX* and *PrmA* protein sequences.** BLAST searches of the *M. silvestris* genome<sup>23</sup> identified two putative SDIMO-encoding gene clusters: Msil\_1262–Msil\_1270 and Msil\_1651–Msil\_1646. The amino acid sequences of the putative  $\alpha$ -subunits, encoded by Msil1262 and Msil1651, designated *mmoX* and *prmA* here, were combined with selected *MmoX* sequences from methanotrophs and related SDIMO sequences from non-methanotrophs and aligned using the Clustal program in MEGA 5 (ref. 32). Positions containing gaps or missing data were eliminated, and 356 amino acid positions in the final data set were used to construct a tree using the maximum likelihood method. Bootstrap values were based on 500 replications.

**DNA and RNA methods.** DNA was extracted from *M. silvestris* and manipulated using standard methods<sup>33</sup>. To visualize transcription from the *mmoX* and *prmA* promoters, the vector pMHA203 (ref. 21), which contains the *mmoX* promoter fused to the GFP gene, was modified by replacement of the *mmoX* promoter sequence with the *prmA* promoter sequence. The primers PrPf (5'-ACTCAATTGTCCGT TCCGTAAAGCCTCTC-3') and PrPr (5'-CGGCCGGCTGAGCTCCGCTAC GC-3') were used to amplify the 1,112 base pair (bp) promoter-containing region extending upstream from 44 bp 5' of the predicted start codon of *prmA*. (The MunI and SacI sites are underlined.) The PCR product was cloned into pCR2.1 TOPO (Invitrogen) and excised by digestion with MunI and SacI. Following EcoRI/SacI excision of the *mmoX* promoter region from pMHA203, the MunI/SacI fragment

was ligated into these sites, resulting in pAC304. pMHA203 and pAC304 were introduced into *M. silvestris* by electroporation<sup>30</sup>, and transformed cells were selected on kanamycin-containing DNMS plates. Cells were visualized using an Axioscop fluorescence microscope (Zeiss).

Total RNA was isolated using a previously described method<sup>34</sup>. Trace residual DNA was removed by two treatments using QIAGEN RNase-free DNase, each followed by purification using an RNeasy spin column (QIAGEN). Reverse transcription was performed using SuperScript II or SuperScript III (Invitrogen), using 50–1,000 ng total RNA with 200 ng random hexamer primers. 5' RACE was performed using the 2nd Generation 5'/3' RACE kit (Roche), following the manufacturer's instructions, using the gene-specific antisense primers 51Ra1 (5'-TCGT CGTCGCATAGCACTTG-3'), 51Ra2 (5'-ATCGTGGAAATGGCGGAACCTC-3') and 51Ra3 (5'-TCTTCTGCATCGGAAAGTACG-3').

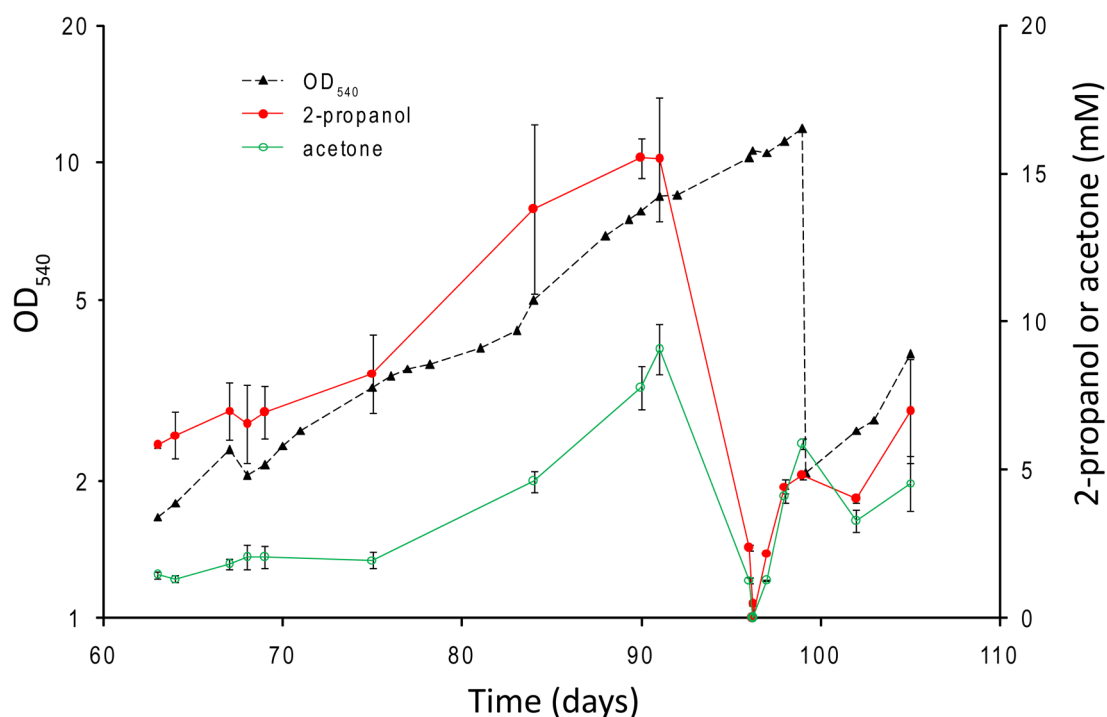
**Mutagenesis.** Mutant strains were constructed as previously described<sup>30</sup>. For disruption of the sMMO, DNA fragments A and B (upstream and downstream of *mmoX*) were amplified by PCR using the primers 1262Af (5'-GATCGAGCTCCGACAC GGAAACAACCTATC-3') and 1262Ar (5'-GATCAGCGTTCGTCGCGGT GCTTAATGC-3'), and 1262Bf (5'-GATCCAAATTGTCGCGATCCGCTCGC AG-3') and 1262Br (5'-GATCGAATTCGATCGAGCGCACAGCTCC-3') (restriction sites underlined). The products were cloned into pCR2.1 TOPO that had been excised with SacI/MluI and MunI/EcoRI and then ligated sequentially into pCM184 (ref. 35) that had been digested with the same enzymes. The linear SacI fragment was introduced into *M. silvestris* by electroporation, and deletion strains were selected on plates containing kanamycin. The kanamycin cassette was subsequently removed by expression of the Cre recombinase following introduction of plasmid pCM157 (ref. 35). The same procedure was followed for the deletion of *prmA*, using the primers 1651Af (5'-GATCGAGCTCTAGTCGGCTACGGCTATTATGG-3') and 1651Ar (5'-GAGAACGCGTGGCGCCTAACGAACCTTCTTTG-3'), and 1651Bf (5'-GATCGGTACCTCATGGGAGCGGATGGATTG-3') and 1651Br (5'-GTC CGCTGACGGTGACTTTG-3'), except that the B fragment was excised from pCR2.1 TOPO using KpnI/EcoRI and ligated into these sites of pCM184. The intended gene deletions were verified by PCR using the primers 1262Tf (5'-CCAGTTCATTTCGTAAGAC-3') and 1262Tr (5'-GTATTCGCTGAACAGCAAGG-3'), and 1651Tf (5'-AAGGCCGCGTCCGATACAAG-3') and 1651Tr (5'-CAGAAC AAATCGGCCCTGGGTCC-3'), which are located outside the cloned regions, and by sequencing.

**Proteomic analyses.** Cells were lysed for proteomic analysis by three passages through a French pressure cell press (American Instrument Company) at 110 MPa (on ice). Cell debris was removed by centrifugation (10,000 g, 15 min, 4 °C). Proteins were separated by SDS-PAGE, and bands of interest were excised from the gels for the identification of polypeptides by the Biological Mass Spectrometry and Proteomics Facility in the School of Life Sciences, University of Warwick. Coomassie-blue-stained gel pieces were processed and tryptically digested using the manufacturer's recommended protocol on the MassPREP robotic protein handling system. The extracted peptides from each sample were analysed by nano liquid-chromatography electrospray-ionization tandem mass spectrometry (LC-ESI-MS/MS) using Nano Acquity/Q-ToF of Ultima Global instrumentation (Waters) with a 45-min liquid chromatography gradient. All MS and MS/MS data were corrected for mass drift using reference data collected from human [Glu<sup>1</sup>]-fibrinopeptide B (catalogue F3261, Sigma) sampled each minute of data collection. The data were used to interrogate an *M. silvestris* database appended with the common Repository of Adventitious Proteins sequences (<http://www.thegpm.org/cRAP/index.html>) using the ProteinLynx Global Server v2.4, as previously described<sup>36</sup>. Polypeptides were identified based on a minimum of two peptides.

**Activity assays.** A Clark-type oxygen electrode equipped with an OXY040A cell connected to a Digital Model 10 controller (Rank Brothers) was used to detect substrate-induced oxygen consumption by whole cells, in a 3 ml working volume. Between 1 mg and 5 mg dry weight of cell suspension was added to 3 ml oxygenated 40 mM phosphate buffer (pH 5.5) in the instrument cell maintained at 25 °C using a circulating water bath. The instrument was operated and calibrated by comparison with air-saturated water as described previously<sup>37</sup>. The substrate (7.5–15,000 nmol) was added, and the substrate-induced rate was calculated by subtracting the endogenous rate from the rate following the addition of substrate. Gaseous substrates were prepared as saturated aqueous solutions in 120 ml serum vials containing 25 ml water flushed with at least ten volumes of the substrate gas, and the concentration was calculated using the Henry's Law constants<sup>22</sup>. Quinoprotein alcohol dehydrogenase was assayed using the artificial electron acceptor phenazine methosulphate (PMS) coupled to reduction of dichlorophenolindophenol (DCPIP), as described previously<sup>38</sup>. Reactions (1 ml) contained 100 mM Tris buffer (pH 9.0), 1 mM PMS, 0.08 mM DCPIP, 15 mM NH<sub>4</sub>Cl, 20 µg protein and 10 mM substrate. Reactions were initiated with the addition of ammonium and followed spectrophotometrically at 600 nm.

**Statistical analysis.** The differences in growth rate, culture density and carbon conversion efficiency shown in Fig. 1 were evaluated using single factor analysis of variance (ANOVA) followed by Tukey's honest significant difference (HSD) method. To identify significant differences in Fig. 3, Student's *t*-test (two-tailed, two samples, assuming equal variance) was applied to the two growth conditions, using Microsoft Excel.

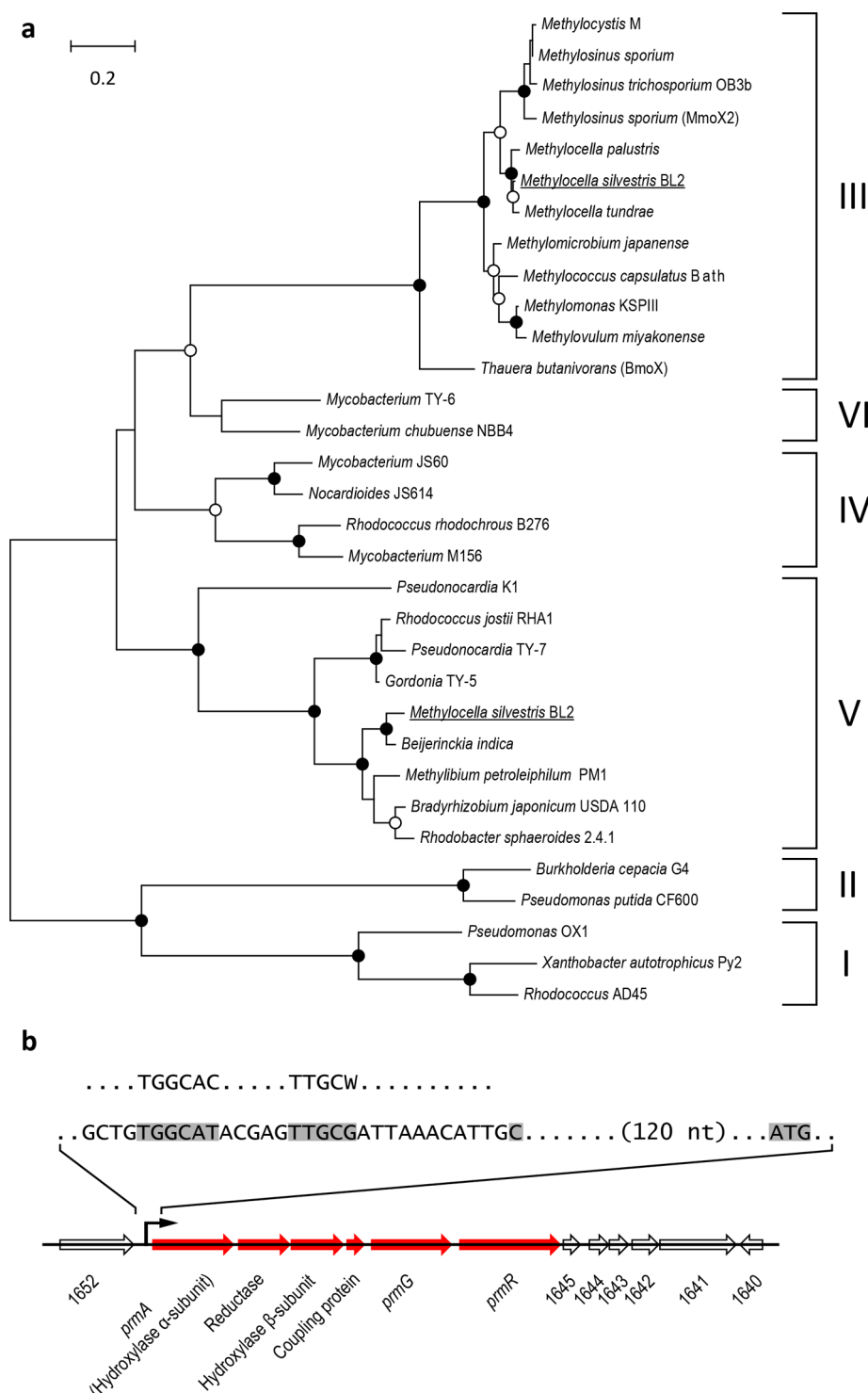
31. Guerlava, P., Izac, V. & Tholozan, J.-L. Comparison of different methods of cell lysis and protein measurements in *Clostridium perfringens*: application to the cell volume determination. *Curr. Microbiol.* **36**, 131–135 (1998).
32. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).
33. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 3rd edn (Cold Spring Harbor Laboratory Press, 2001).
34. Gilbert, B. *et al.* Molecular analysis of the *pmo* (particulate methane monooxygenase) operons from two type II methanotrophs. *Appl. Environ. Microbiol.* **66**, 966–975 (2000).
35. Marx, C. J. & Lidstrom, M. E. Broad-host-range *cre-lox* system for antibiotic marker recycling in Gram-negative bacteria. *Biotechniques* **33**, 1062–1067 (2002).
36. Patel, V. J. *et al.* A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *J. Proteome Res.* **8**, 3752–3759 (2009).
37. Green, M. J. & Hill, H. A. O. Chemistry of dioxygen. *Methods Enzymol.* **105**, 3–22 (1984).
38. Anthony, C. & Zatman, L. J. The microbial oxidation of methanol. 2. The methanol-oxidizing enzyme of *Pseudomonas* sp. M 27. *Biochem. J.* **92**, 614–621 (1964).
39. Leahy, J. G., Batchelor, P. J. & Morcomb, S. M. Evolution of the soluble diiron monooxygenases. *FEMS Microbiol. Rev.* **27**, 449–479 (2003).
40. Coleman, N. V., Bui, N. B. & Holmes, A. J. Soluble di-iron monooxygenase gene diversity in soils, sediments and ethene enrichments. *Environ. Microbiol.* **8**, 1228–1239 (2006).
41. Barrios, H., Valderrama, B. & Morett, E. Compilation and analysis of  $\sigma^{54}$ -dependent promoter sequences. *Nucleic Acids Res.* **27**, 4305–4313 (1999).



**Extended Data Figure 1 | *Methylocella silvestris* grown on propane (20% (v/v) propane in air) in a fermenter (period between 62 and 106 days shown).** The 2-propanol and acetone in the culture medium reached approximately 16 mM and 9 mM, respectively. When the propane supply was shut off (days 92–96), the 2-propanol and acetone concentrations decreased to nearly zero, whereas the culture density continued to increase without interruption. Following the resumption of the propane supply, accumulation of

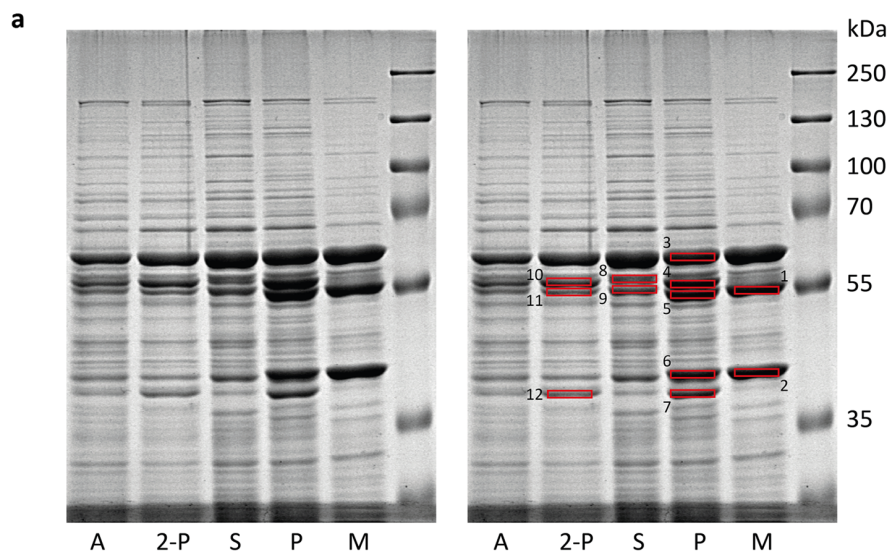
2-propanol and acetone was observed once again. The cells and medium were removed, and fresh medium was added on day 99. These data demonstrate that 2-propanol and acetone result from propane oxidation and that cells growing on propane in the presence of these intermediates can metabolize them without an appreciable lag phase. The 2-propanol and acetone concentrations show the mean  $\pm$  s.d. of triplicate measurements.





**Extended Data Figure 2 | Phylogeny of SDIMOs.** **a**, Phylogenetic relationships between the two *M. silvestris* SDIMOs (underlined) and other representative enzymes. The tree, constructed using the maximum likelihood method, is based on an alignment of amino acid sequences of the  $\alpha$ -subunit of the hydroxylases. The sequences were aligned using Clustal; positions containing gaps or missing data were eliminated; and the tree was constructed with a final data set of 356 amino acids using MEGA 5 (ref. 32). Bootstrap values (based on 500 replications) greater than 95% are shown as filled circles at nodes, and those between 75% and 95% as open circles. The SDIMO subgroups<sup>39,40</sup> are indicated on the right of the figure. GenBank accession numbers (in order from the top): AAC45289.1, ABD46892.1, ZP\_06887019.1, ABD46898.1, CAD30366.1, YP\_002361593.1, CAD88243.1, BAE86875.1,

YP\_113659.1, BAA84751.1, BAJ17645.1, AAM19727.1, BAF34294.1, ACZ56324.1, AAO48576.1, YP\_919254.1, BAA07114.1, AAS19484.1, CAC10506.1, YP\_700435.1, BAF34308.1, BAD03956.2, YP\_002361961.1, YP\_001834443.1, YP\_001020147.1, NP\_770317.1, YP\_352924.1, AAL50373.1, P19732.1, AAT40431.1, YP\_001409304.1 and CAB55825.1. **b**, The propane monooxygenase gene cluster (shown in red). The structural genes (hydroxylase  $\alpha$ -subunit, reductase, hydroxylase  $\beta$ -subunit and coupling protein) are followed by those encoding a putative chaperone *prmG* (ACK50595.1) and regulatory protein *prmR* (ACK50594.1), with homology to *mmoG* and *mmoR* of sMMO. The putative promoter sequence is shown above in relation to the ATG start codon, together with the consensus sequence of  $\sigma^{54}$  promoters as described previously<sup>41</sup>.

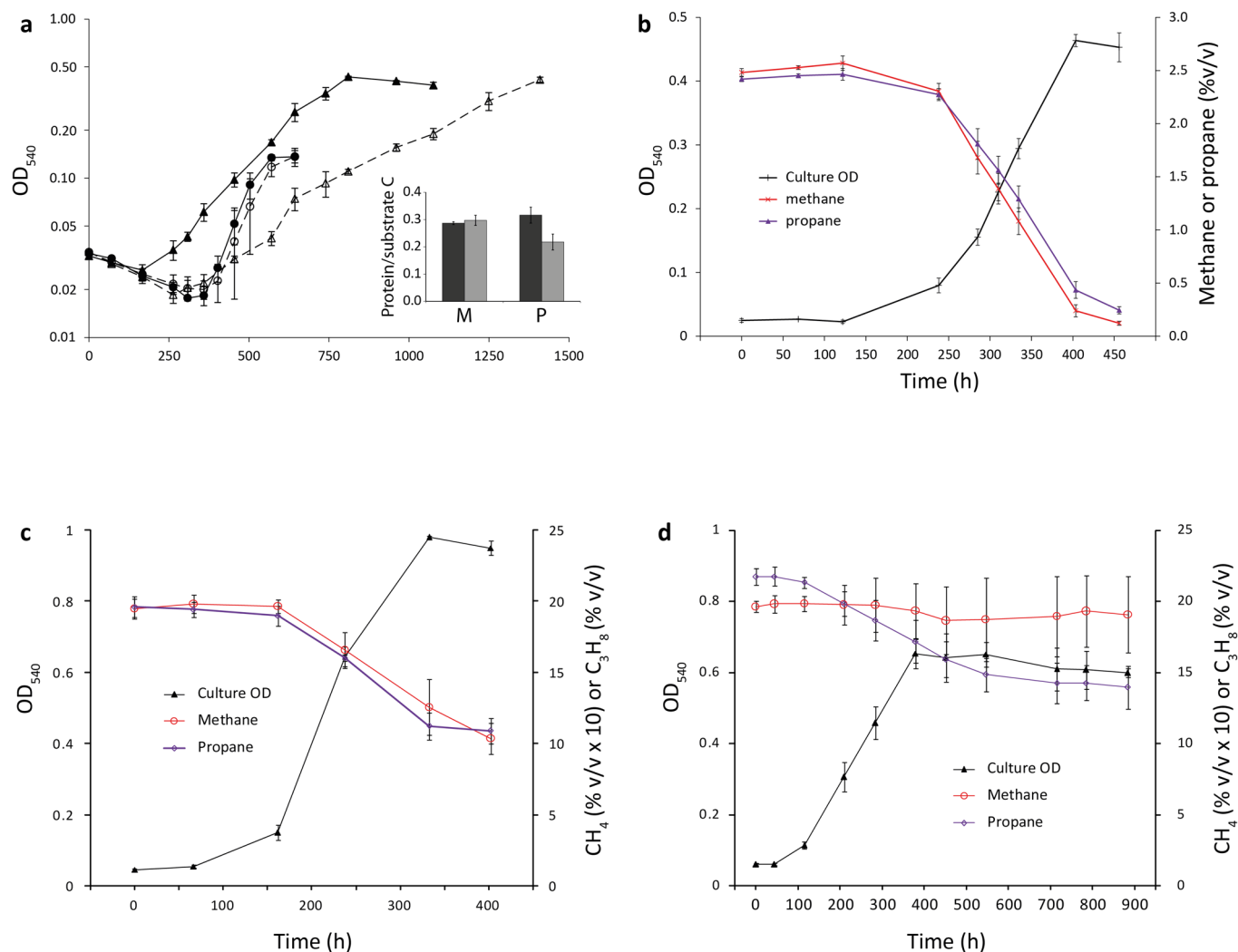


**b**

Band	Growth condition	Gene i.d.	Annotation	Peptides	Total peptides	MM (kDa)	Band	Growth condition	Gene i.d.	Annotation	Peptides	Total peptides	MM (kDa)
1	Methane	1262	sMMO $\alpha$ -subunit	22	28	59.8	7	Propane	1649	PrMO $\beta$ -subunit	11	22	40.2
		1263	sMMO $\beta$ -subunit	4		44.9			1263	sMMO $\beta$ -subunit	5		44.9
		2342	Aldehyde DH	2		55.6			2400	Formylmethanofuran DH	4		39.2
									1265	sMMO $\gamma$ -subunit	2		19.5
2	Methane	1263	sMMO $\beta$ -subunit	25	27	44.9	8	Succinate	0795	Chaperonin	20	119	57.6
		1262	sMMO $\alpha$ -subunit	2		59.8			3763	Hydrogenase	16		59.9
3	Propane	0471	Methanol DH $\alpha$ -subunit	33	33	68.5			3631	Nitrogenase s/u	12		58.2
									3630	Nitrogenase s/u	11		58.2
4	Propane	1651	PrMO $\alpha$ -subunit	18	59	64.2	9	Succinate	2342	Aldehyde DH	17	46	55.6
		1641	GMC oxidoreductase	9		56.7			3881	Aldehyde DH	8		53.3
		0795	Chaperonin GroEL	7		57.6			2301	Unknown function	4		55.8
		1375	PEP carboxykinase	6		58.2			1608	FeS assembly protein	3		54.3
		1647	Chaperonin (PrMO cluster)	3		57.3	10	2-propanol	1651	PrMO $\alpha$ -subunit	29	86	64.2
		2506	Dihydrolipoyl DH	3		49.4			1641	GMC oxidoreductase	16		56.7
		0162	Fumarase	3		58.6			1647	Chaperonin (PrMO cluster)	12		57.3
		1262	sMMO $\alpha$ -subunit	2		59.8			1375	PEP carboxykinase	8		58.2
5	Propane	1262	sMMO $\alpha$ -subunit	20	48	59.8	11	2-propanol	2342	Aldehyde DH	19	57	55.6
		2342	Aldehyde DH	12		55.6			1651	PrMO $\alpha$ -subunit	12		64.2
		1651	PrMO $\alpha$ -subunit	5		64.2			3881	Aldehyde DH	10		53.3
		0410	Beta-lactamase	5		52.4			2810	Phosphoglucosmutase	5		54.0
		3881	Aldehyde DH	3		53.3			1641	GMC oxidoreductase	4		56.7
		1263	sMMO $\beta$ -subunit	3		44.9	12	2-propanol	1649	PrMO $\beta$ -subunit	16	31	40.2
6	Propane	1263	sMMO $\beta$ -subunit	23	49	44.9			2400	Formylmethanofuran DH	4		39.2
		2007	Urea transporter	9		43.4			1194	Fructose bisphosphate aldolase	3		39.1
		3698	NADH flavin oxidase	5		39.2			2991	RNA polymerase $\alpha$ -subunit	3		37.1
		1267	sMMO reductase	5		38.5							
		2996	Acetyl-CoA acetyltransferase	3		40.2							
		1262	sMMO $\alpha$ -subunit	2		59.8							

**Extended Data Figure 3 | *M. silvestris* polypeptide profiles.** **a**, An SDS-PAGE gel loaded with soluble extract from wild-type cells grown on methane (M), propane (P), succinate (S), 2-propanol (2-P) or acetone (A). Prominent bands (identified with boxes in the right-hand photograph of the same gel) evident following growth on methane or propane, but not on succinate, were excised from the gel and analysed by mass spectrometry. The polypeptide identifications are shown in **b**. The data show that both sMMO and PrMO were expressed during growth on propane but that PrMO subunits were not expressed at a high level during growth on methane. PrMO subunits, but not sMMO subunits, were expressed during growth on 2-propanol. In addition, gel-free analysis of the complete soluble proteome<sup>26</sup> did not result in the detection of PrMO polypeptides in succinate-grown or methane-grown cell

extracts. **b**, Polypeptide identifications of the gel bands shown in **a**. For each band, the four most abundant polypeptides are shown, except where fewer than four were detected. In addition, all sMMO-related and PrMO-related polypeptides identified (irrespective of the number of peptides used for identification) are included. Otherwise, polypeptides identified with at least three peptides are included (except for the succinate lane; four peptides). Other identified polypeptides are not shown. The number of peptides used for the identification of each polypeptide is shown. The total number of peptides detected from all of the polypeptides identified in each band is shown for comparison. DH, dehydrogenase; MM, theoretical molecular mass; PEP, phosphoenolpyruvate.



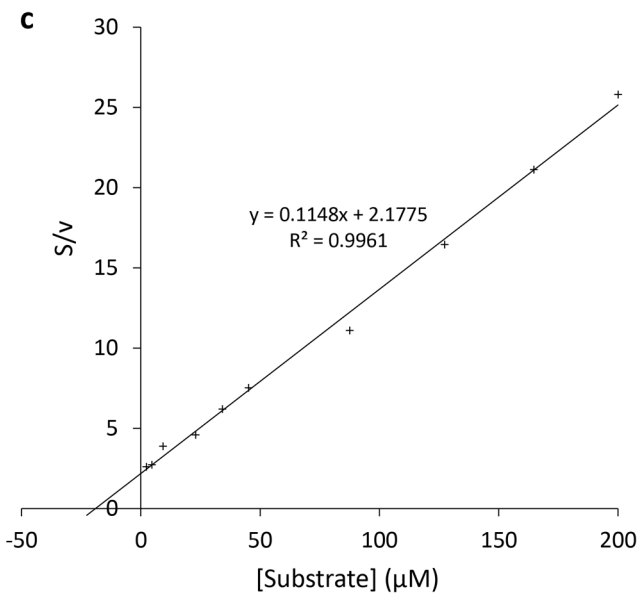
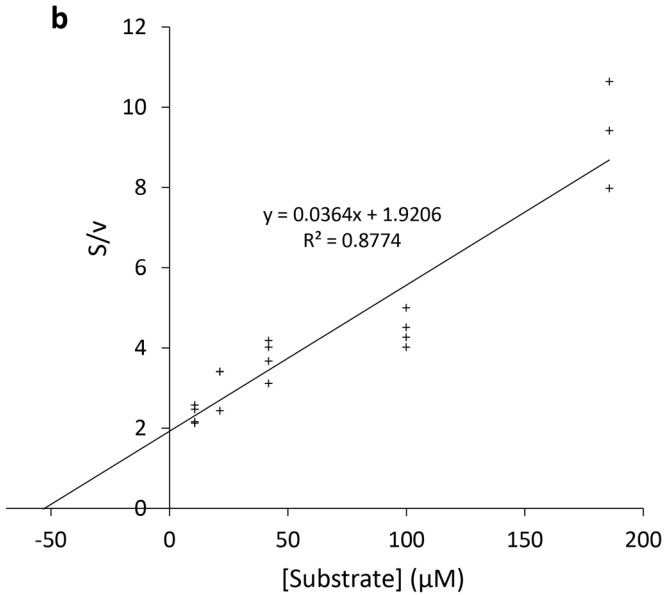
**Extended Data Figure 4 | Growth of *M. silvestris* strains.** **a**, Growth of wild-type *M. silvestris* (solid lines, filled symbols) and the  $\Delta$ PrmA strain (dashed lines, open symbols) on methane (circles) or propane (triangles) (2.5% (v/v)). The inset shows the substrate carbon conversion efficiency (CCE) (mg of cellular protein formed per mg of substrate carbon consumed) for the wild-type *M. silvestris* (dark grey) and the  $\Delta$ PrmA strain (light grey) during growth on methane (M) or propane (P). Data are the mean  $\pm$  s.d. for three (methane) or two (propane) vials. **b**, Growth and substrate consumption of

wild-type *M. silvestris* on a mixture of methane (red, crosses) and propane (purple, triangles) (2.5% (v/v) each). The inoculum for the cultures was grown on methane and propane (2.5% (v/v) each). Data points show the mean  $\pm$  s.d. of triplicate vials. **c**, **d**, Growth of and consumption of methane and propane by wild-type *M. silvestris* (**c**) and the  $\Delta$ MmoX strain (**d**) when supplied with an approximately 1:10 (v/v) ratio of gases, showing that PrMO did not oxidize appreciable amounts of methane. Methane concentrations are shown as % (v/v)  $\times 10$  on the secondary (right) y axes. Data are the mean  $\pm$  s.d. of triplicate vials.



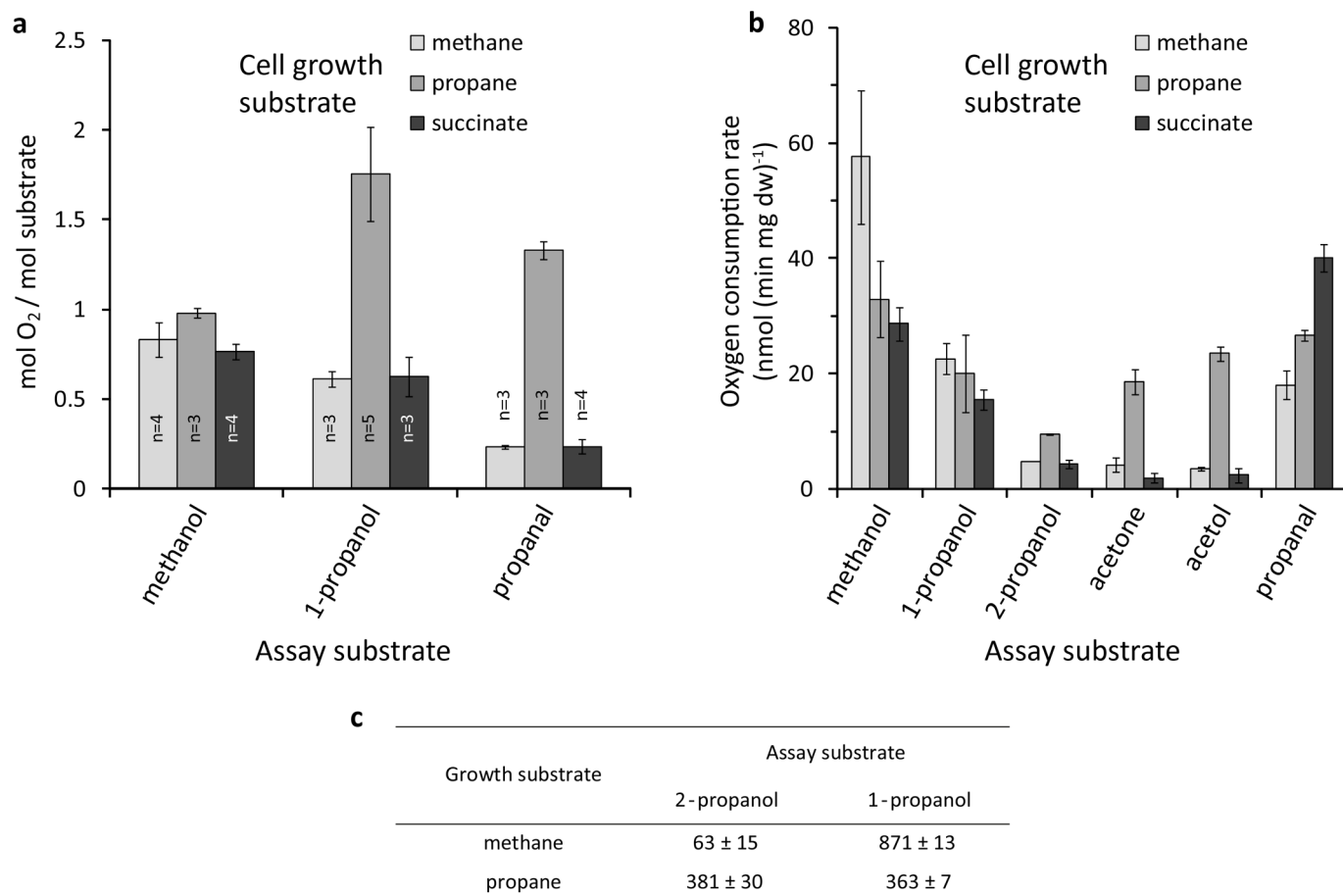
**a**

Substrate for oxidation	Growth substrate		
	Methane	Propane	Succinate
Methane	49.0 ± 1.2	21.1 ± 2.1	Not detected
Propane	2.5 ± 0.5	5.9 ± 0.8	Not detected



**Extended Data Figure 5 | Methane and propane oxidation kinetics.**  
**a**, Maximum methane-induced and propane-induced specific oxygen consumption rate ( $\text{nmol min}^{-1}$  per mg dry weight) of whole *M. silvestris* cells grown on methane, propane or succinate, as determined by oxygen electrode studies, with approximately 200  $\mu\text{M}$  oxidation substrate. Data are the

mean  $\pm$  s.d. of three measurements. **b, c**, Kinetics of methane-induced (**b**) or propane-induced (**c**) oxygen consumption in whole cells grown on that substrate. The Hanes–Woolf plots show substrate concentration ( $S$ ) divided by oxygen consumption rate ( $v$ ) as a function of  $S$ , and the trend line cuts the  $x$  axis at  $-K_m$ .

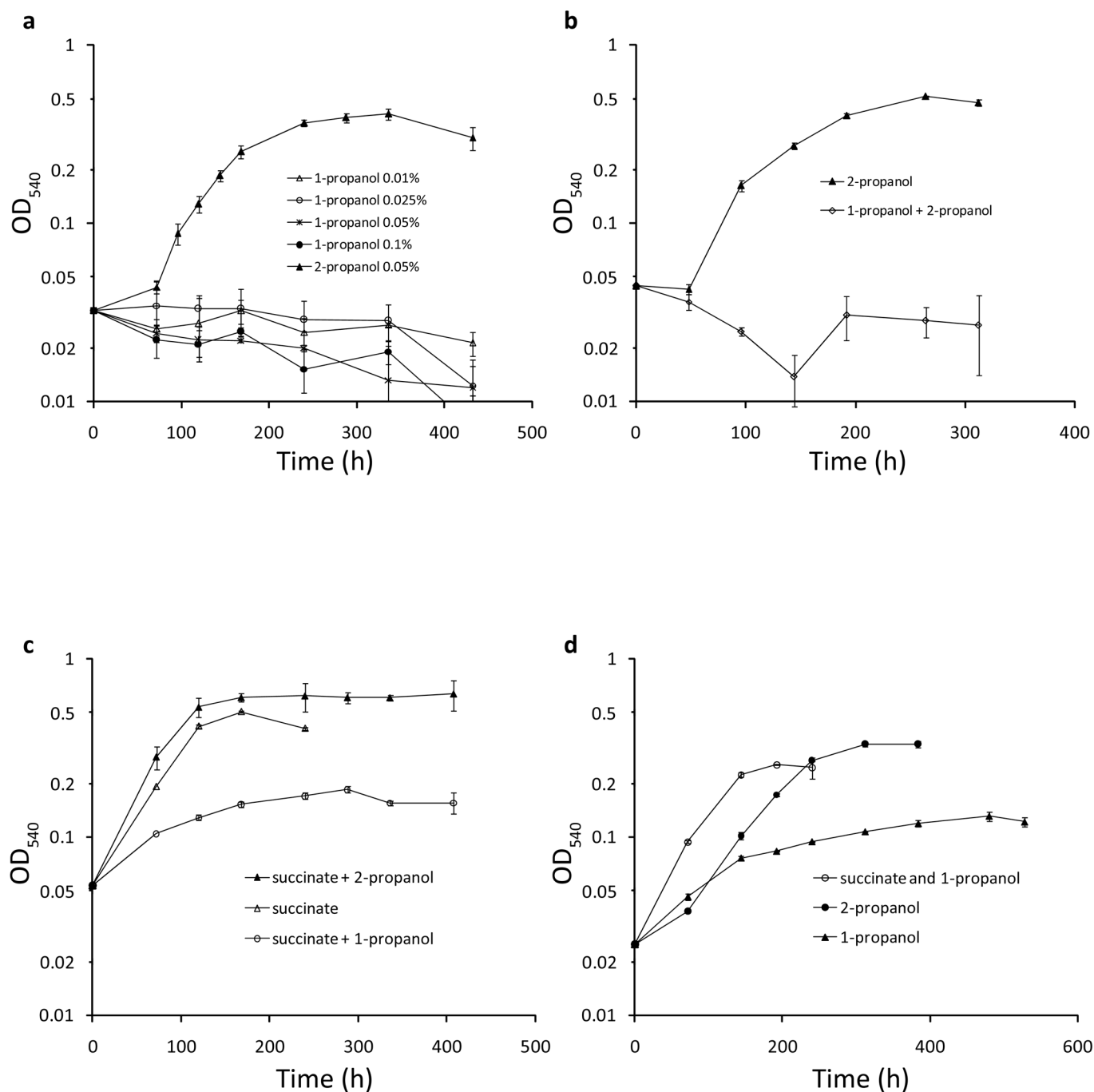


#### Extended Data Figure 6 | Wild-type *M. silvestris* activity assays.

**a**, Stoichiometry of whole-cell oxygen consumption in response to addition of 100–250 nmol methanol, 1-propanol or propanal. Cells from all growth conditions consumed approximately equimolar amounts of oxygen when methanol was added; however, when 1-propanol or propanal replaced methanol, the amount of oxygen consumed by propane-grown cells increased to threefold (1-propanol) or fivefold (propanal) that of methane-grown or succinate-grown cells. These data suggest that the oxidation of terminal intermediates proceeds further in propane-grown cells, although enzymes that can oxidize 1-propanol and propanal may be present in all cell types, as shown in **b** and **c**. Data are the mean ± s.d. of the number of measurements shown.

**b**, Oxygen uptake rates of *M. silvestris* cells grown on methane, propane or succinate in response to the addition of the substrates shown. High rates of oxygen consumption were recorded in response to methanol, 1-propanol and propanal in cells grown on either methane, propane or succinate. However, the oxygen consumption rates in response to 2-propanol, acetone and acetol were

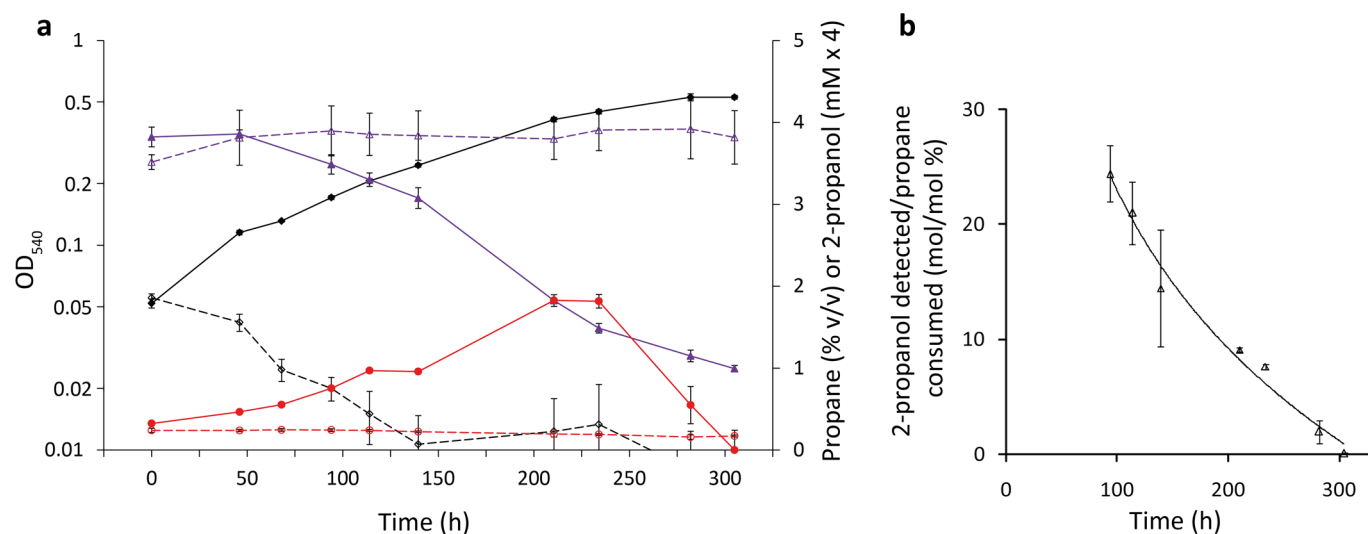
at least twofold, fourfold or sevenfold higher, respectively, in propane-grown cells than in methane-grown or succinate-grown cells, demonstrating that the ability to oxidize intermediates of the sub-terminal oxidation pathway was induced during growth on propane. Substrates were used at a final concentration of 5 mM. Data are the mean ± s.d. of three measurements (except for methane/methanol,  $n = 7$ ). **c**, Quinoprotein alcohol dehydrogenase activity (mean ± s.d. of three measurements) assayed as DCPIP reduction (nmol min<sup>-1</sup> per mg protein) in soluble extract from methane-grown or propane-grown cells. The activity in cell extracts from both growth conditions was high when 1-propanol was the substrate (probably as a result of the constitutive expression of methanol dehydrogenase); however, with 2-propanol as the substrate, the activity was sixfold higher in the extract from cells grown on propane than cells grown on methane. By contrast, the rates of NAD<sup>+</sup>-linked or NADP<sup>+</sup>-linked 1-propanol or 2-propanol dehydrogenase activity were less than 10 nmol min<sup>-1</sup> per mg protein in cell extracts from each growth condition (data not shown).



**Extended Data Figure 7 | 1-Propanol-metabolizing ability was induced in propane-grown cells.** **a**, When vials were inoculated with methane-grown cells, growth was possible on 2-propanol but not on 1-propanol at any concentration tested. **b**, In addition, 1-propanol completely inhibited growth on 2-propanol. **c**, When using succinate-grown inoculum, 1-propanol also greatly inhibited growth on succinate. **d**, However, when using an inoculum grown on propane, limited growth occurred on 1-propanol, and 1-propanol did

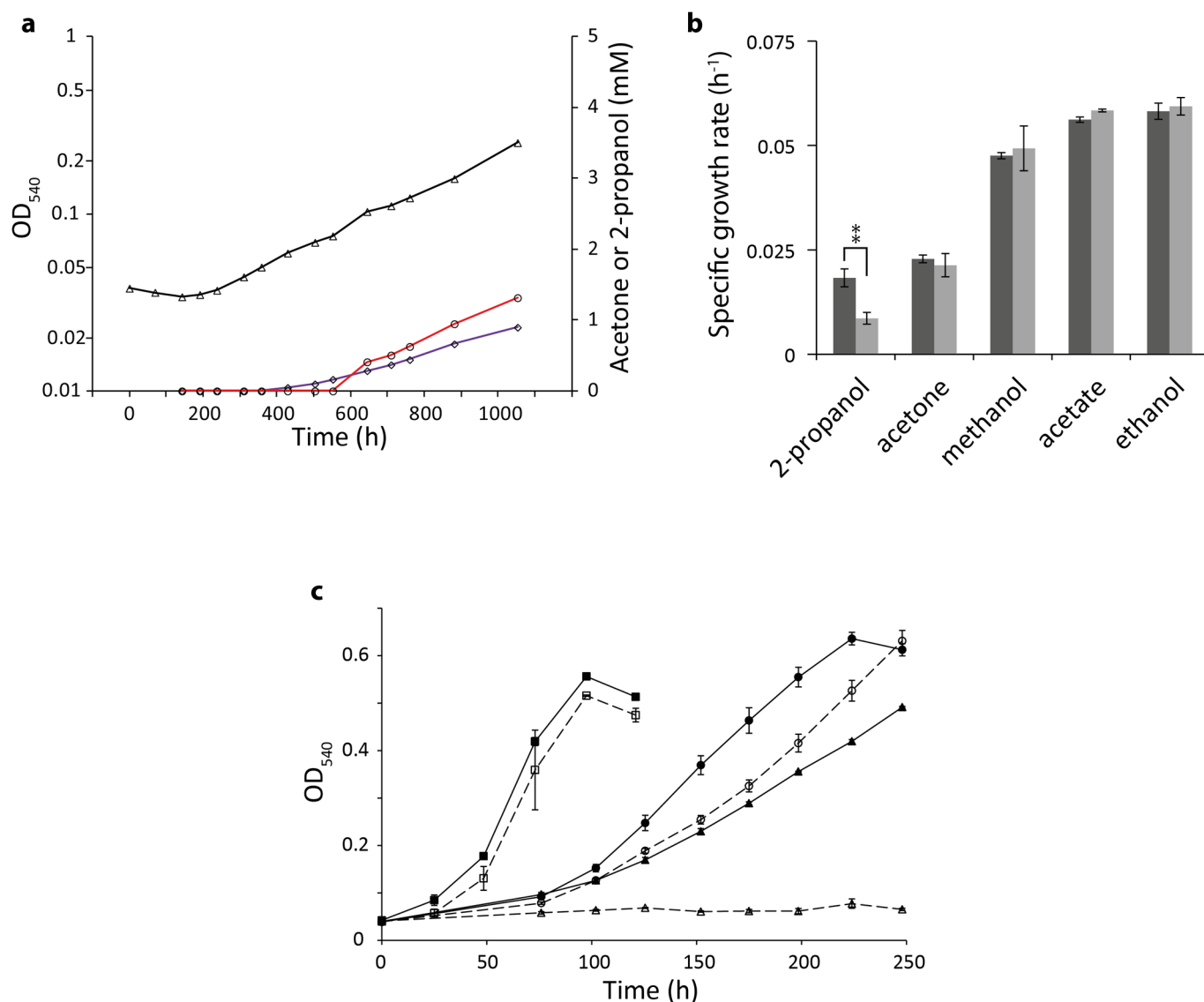
not inhibit growth on succinate, suggesting that 1-propanol-metabolizing potential was induced in cells grown on propane and could be maintained during growth on, or in the presence of, 1-propanol. The concentrations used were 0.05% (v/v) 1-propanol, 0.05% (v/v) 2-propanol, 3 mM succinate (except **d**, 5 mM), or as indicated. Data are the mean  $\pm$  s.d. of duplicate (**a**, **c** **d**) or triplicate (**b**) vials.





**Extended Data Figure 8 | Detection of oxidation products in wild-type *M. silvestris* cultures.** **a**, During growth on 4% (v/v) propane, 2-propanol was detected in the culture medium during the mid-exponential and late exponential phase, before declining from a maximum of approximately 0.5 mM at 210 h to below the limit of detection at stationary phase (300 h). The culture density (OD<sub>540</sub>) is shown in black; propane concentration, in purple; and 2-propanol concentration, in red; with solid lines and filled symbols. Control vials, which contained cells killed by autoclaving, are shown as dotted lines and open symbols. The 2-propanol concentration is shown as  $\times 4$  on the secondary

(right) y axis. Data are the mean  $\pm$  s.d. of triplicate vials. **b**, For the cultures shown in **a**, the amount of 2-propanol present at each time point (including the amount previously removed during sampling) is expressed as a percentage of the propane consumed. At 94 h, 25% of the propane consumed could be accounted for as 2-propanol in the growth medium. This value therefore represents the minimum percentage of propane that is oxidized to 2-propanol, because no allowance has been made for the consumption of 2-propanol by the cultures. Data are calculated from the mean  $\pm$  s.d. of triplicate vials.



**Extended Data Figure 9 | Oxidation products of the  $\Delta$ PrmA strain and growth rates of the  $\Delta$ PrmA, wild-type and inhibited wild-type strains.**

**a**, Culture density (black lines, triangles) and 2-propanol (red lines, circles) and acetone (purple lines, diamonds) concentrations during growth of the  $\Delta$ PrmA strain on 4% (v/v) propane. This strain accumulated 1.3 mM 2-propanol, but neither 2-propanol nor acetone were detected during growth of the  $\Delta$ MmoX strain on 20% (v/v) propane (data not shown), suggesting that 2-propanol is one (or the major) product of propane oxidation by sMMO. **b**, Growth rate of wild-type *M. silvestris* (dark grey) and the  $\Delta$ PrmA strain (light grey) on the substrates shown. Disruption of PrMO decreased growth on 2-propanol compared with the wild-type, although growth on acetone, acetate, methanol or

ethanol was not significantly affected, implicating PrMO in 2-propanol metabolism. Data are the mean  $\pm$  s.d. of three vials (except 2-propanol,  $n = 6$  ( $\Delta$ PrmA strain) or  $n = 5$  (wild-type)). Significance was determined by Student's *t*-test (\*\*,  $P < 0.01$ ). **c**, Growth of wild-type *M. silvestris* on acetone (squares), 2-propanol (circles) or propane (10% (v/v)) (triangles) either without an inhibitor (filled symbols, solid lines) or with acetylene (2% (v/v)) as an inhibitor (open symbols, dashed lines). Data are the mean  $\pm$  s.d. of triplicate vials. For 2-propanol, the specific growth rates in the early exponential phase were  $0.020 \pm 0.0002 \text{ h}^{-1}$  and  $0.018 \pm 0.0001 \text{ h}^{-1}$  (mean  $\pm$  s.d.) for uninhibited or inhibited cultures, respectively.

Extended Data Table 1 | Substrate utilization by *Methylocella silvestris*

Substrate	Concentration	Growth
Formate	1 mM	-
Formate	5 mM, 40 mM	+
Formate	25 mM	++
H <sub>2</sub> /CO <sub>2</sub>	20% O <sub>2</sub> , 0.5 to 10% CO <sub>2</sub> , balance H <sub>2</sub>	-
Methane	2.5% to 20%	++
Methanol	0.05%	++
Methylamine	5 mM	++
Acetate	5 mM	++
Ethane	20%	+
Ethene	10%	-
Ethanol	0.05%	++
Glycine	5 mM	-
Glycollate	5 mM	-
Glyoxylate	1 mM, 5 mM	-
Oxalate	1 mM, 5 mM	-
Urea	5 mM	-
1,2-propanediol	0.05%	++
1-propanol	0.01 to 0.1%	.*
2-propanol	0.05%	++
Acetol	0.05%	++
Acetone	0.05%	++
Glycerol	5 mM	++
Methyl acetate	0.05%	+
Propane	2.5% to 20%	++
Propene	10%	-
Propionate	5 mM	++
Propionate	20 mM	-
Pyruvate	5 mM	++
Butane	20%	-
Trans-2-butene	10%	-
Malate	5 mM	++
Succinate	5 mM	++
Tetrahydrofuran	0.05%	+
Pentane	0.1%	-
Gluconate	10 mM	++
Pentadecane	0.1%	-

Growth is indicated as: ++, growth to OD<sub>540</sub> > 0.25; +, growth to OD<sub>540</sub> ≥ 0.08–0.25; –, growth to OD<sub>540</sub> < 0.08. (Concentrations in % are given as v/v.)

\* Growth on 1-propanol was dependent on the inoculum used for the cultures.



# Focused specificity of intestinal T<sub>H</sub>17 cells towards commensal bacterial antigens

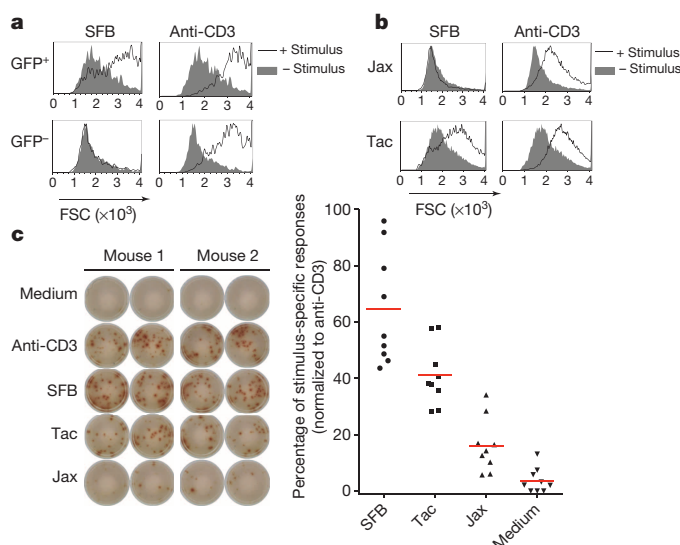
Yi Yang<sup>1</sup>, Miriam B. Torchinsky<sup>1</sup>, Michael Gobert<sup>1</sup>, Huizhong Xiong<sup>1</sup>, Mo Xu<sup>1</sup>, Jonathan L. Linehan<sup>2†</sup>, Francis Alonzo<sup>3</sup>, Charles Ng<sup>1</sup>, Alessandra Chen<sup>1</sup>, Xiyao Lin<sup>1</sup>, Andrew Sczesnak<sup>1</sup>, Jia-Jun Liao<sup>1</sup>, Victor J. Torres<sup>3</sup>, Marc K. Jenkins<sup>2</sup>, Juan J. Lafaille<sup>1</sup> & Dan R. Littman<sup>1,4</sup>

T-helper-17 (T<sub>H</sub>17) cells have critical roles in mucosal defence and in autoimmune disease pathogenesis<sup>1–3</sup>. They are most abundant in the small intestine lamina propria, where their presence requires colonization of mice with microbiota<sup>4–7</sup>. Segmented filamentous bacteria (SFB) are sufficient to induce T<sub>H</sub>17 cells and to promote T<sub>H</sub>17-dependent autoimmune disease in animal models<sup>8–14</sup>. However, the specificity of T<sub>H</sub>17 cells, the mechanism of their induction by distinct bacteria, and the means by which they foster tissue-specific inflammation remain unknown. Here we show that the T-cell antigen receptor (TCR) repertoire of intestinal T<sub>H</sub>17 cells in SFB-colonized mice has minimal overlap with that of other intestinal CD4<sup>+</sup> T cells and that most T<sub>H</sub>17 cells, but not other T cells, recognize antigens encoded by SFB. T cells with antigen receptors specific for SFB-encoded peptides differentiated into RORγt-expressing T<sub>H</sub>17 cells, even if SFB-colonized mice also harboured a strong T<sub>H</sub>1 cell inducer, *Listeria monocytogenes*, in their intestine. The match of T-cell effector function with antigen specificity is thus determined by the type of bacteria that produce the antigen. These findings have significant implications for understanding how commensal microbiota contribute to organ-specific autoimmunity and for developing novel mucosal vaccines.

How SFB induces T<sub>H</sub>17 cells and how these cells contribute to self-reactive pathological responses remain key unanswered questions. A recent study, using mice with monoclonal TCRs, suggested that induction of T<sub>H</sub>17 cells by SFB or other microbiota is independent of cognate antigen recognition<sup>15</sup>. To further evaluate mucosal effector T-cell induction in a physiological setting, we undertook an examination of the repertoire and specificity of naturally arising T<sub>H</sub>17 cells. To facilitate analysing live T<sub>H</sub>17 cells, we used *Il23r<sup>GFP</sup>* reporter mice<sup>16</sup>, as among CD4<sup>+</sup> T cells, only this subset expresses IL-23R. We first asked if small intestine lamina propria (SILP) T<sub>H</sub>17 cells are in general responsive to gut luminal commensal antigens. GFP<sup>+</sup> (T<sub>H</sub>17) and GFP<sup>–</sup> (non-T<sub>H</sub>17) CD4<sup>+</sup> T cells, purified from *Il23r<sup>GFP/+</sup>* C57BL/6 (B6) mice that had been colonized with SFB, were incubated with splenic antigen-presenting cells (APCs) and autoclaved small intestinal luminal content of mice from the Jackson laboratory (Jackson) and Taconic Farms (Taconic). We used the measure of forward scatter (FSC) as a surrogate readout for T-cell activation. Intriguingly, only T<sub>H</sub>17 cells mounted a detectable response to Taconic antigens (Extended Data Fig. 1a). SFB is one of the bacteria unique to Taconic flora<sup>8</sup>. Thus we repeated the assay with faecal material from SFB-monoassociated mice (SFB-mono antigens) and detected a robust response only among GFP<sup>+</sup> cells (Fig. 1a). These cells did not respond to major histocompatibility complex class II (MHCII)-deficient APCs loaded with SFB-mono antigens, indicating that the activation was dependent on antigen presentation (Extended Data Fig. 1b). SFB-mono antigens selectively stimulated total CD4<sup>+</sup> T cells from B6 Taconic mice, but not those from B6 Jackson mice, consistent with

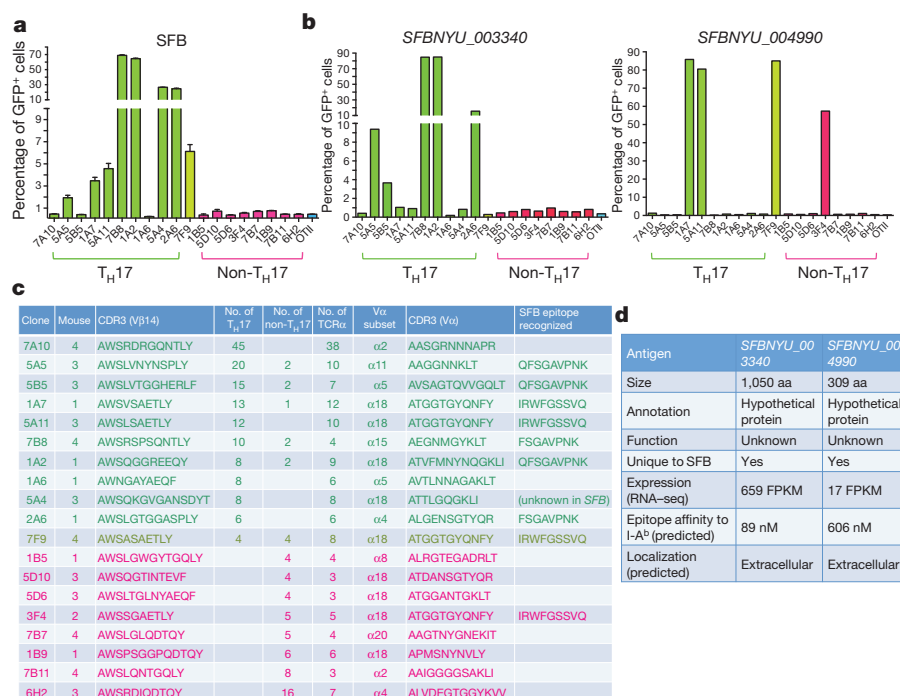
*in vivo* priming of SFB-specific T<sub>H</sub>17 cells (Fig. 1b), and any bystander effect in this assay was negligible (Extended Data Fig. 1c). Next, we used an IL-17A ELISPOT assay to quantify the percentage of T<sub>H</sub>17 cells from SFB-colonized mice responding to commensal antigens. GFP<sup>+</sup> cells had a relatively weak response towards Jackson antigens, but had a robust response towards Taconic antigens. Significantly, SFB mono-associated mouse faecal antigens stimulated over 60% of the T<sub>H</sub>17 cells (Fig. 1c). In contrast, there was no response of T<sub>H</sub>17 cells to faecal material from germ-free mice (data not shown). Thus, the majority of T<sub>H</sub>17 cells in the SILP of SFB-colonized mice react with SFB-derived antigens, whereas a small proportion respond to non-SFB antigen, indicating that most T<sub>H</sub>17 cells are specific for bacteria in the intestinal lumen.

We wished to compare the TCR repertoires of T<sub>H</sub>17 cells and those of non-T<sub>H</sub>17 cells. Using antibodies against a panel of TCR Vβs, we observed a higher proportion of Vβ14<sup>+</sup> T cells in T<sub>H</sub>17 cells than in non-T<sub>H</sub>17 cells from the SILP (Extended Data Fig. 2a, b). This bias was



**Figure 1 | Intestinal T<sub>H</sub>17 cells are specific for SFB- and other microbiota-derived antigens.** **a**, Selective activation of intestinal GFP<sup>+</sup> CD4<sup>+</sup> T cells from *Il23r<sup>GFP/+</sup>* mice by faecal extract from SFB-monoassociated mice. Forward scatter (FSC) was evaluated after 2 days. **b**, Activation of SILP CD4<sup>+</sup> T cells from B6 Taconic mice and B6 Jackson mice with faecal extract from SFB-monoassociated mice. Jax, Jackson mice; Tac, Taconic mice. **c**, IL-17A ELISPOT assay of intestinal GFP<sup>+</sup> CD4<sup>+</sup> T cells from SFB-colonized *Il23r<sup>GFP/+</sup>* mice treated with indicated stimuli. Left, representative ELISPOT images. Right, compilation of results from multiple animals. Each symbol represents cells from a separate animal.

<sup>1</sup>The Kimmel Center for Biology and Medicine of the Skirball Institute, New York University School of Medicine, New York, New York 10016, USA. <sup>2</sup>Department of Microbiology, Center for Immunology, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA. <sup>3</sup>Department of Microbiology, New York University School of Medicine, New York, New York 10016, USA. <sup>4</sup>Howard Hughes Medical Institute, New York University School of Medicine, New York, New York 10016, USA. <sup>†</sup>Present address: Mucosal Immunology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892-0485, USA.



**Figure 2 | Most  $T_H17$  TCR hybridomas recognize SFB-unique proteins.**

**a**, Responses of the TCR hybridomas, prepared from  $T_H17$  and non- $T_H17$  intestinal  $CD4^+$  T cells, to faecal material from SFB-monoassociated mice. **b**, Responses of the TCR hybridomas to *Escherichia coli* clones expressing full-length *SFBNYU\_003340* and *SFBNYU\_004990*. Note that a non- $T_H17$  TCR

hybridoma also responded to the clone expressing *SFBNYU\_004990*. **c**, Summary of the nineteen dominant clonotypic TCR clones. Ten  $T_H17$ -biased clones are highlighted in green, and eight non- $T_H17$ -biased clones are highlighted in red. **d**, Features of the two antigenic proteins of SFB. FPKM, fragments per kilobase of transcript per million mapped reads.

recapitulated when the  $CD4^+$  T cells were stained with antibodies specific for ROR $\gamma$ t and IL-17A, two other characteristic markers of  $T_H17$  cells (Extended Data Fig. 2c). However, intracellular staining for IFN $\gamma$  and FOXP3 indicated no V $\beta$ 14 $^+$  cell bias among  $T_H1$  and T regulatory cells (Extended Data Fig. 2c). To determine if the V $\beta$ 14 enrichment of  $T_H17$  cells is influenced by microbiota, we compared SFB-free B6 Jackson mice with SFB-colonized B6 Taconic mice. The Jackson mice had few ROR $\gamma$ t $^+$   $T_H17$  cells, and there was no enrichment of V $\beta$ 14 $^+$  cells among them. In contrast, Jackson mice cohoused with Taconic mice had increased numbers of lamina propria  $T_H17$  cells, which were enriched for V $\beta$ 14 $^+$  TCRs (Extended Data Fig. 2d), indicating that the  $T_H17$  repertoire is shaped by specific microbiota.

We chose to focus on V $\beta$ 14 $^+$  cells to further elucidate the gut  $CD4^+$  T-cell repertoire. First, we used pyrosequencing to examine the repertoire of V $\beta$ 14 $^+$  SILP  $T_H17$  and non- $T_H17$  cells from SFB-colonized mice. The complementarity-determining region 3 (CDR3) of V $\beta$ 14 was characterized for each cell population from eight *Il23r<sup>GFP/+</sup>* mice. Each sample contained a minimum of several hundred unique CDR3 sequences (Extended Data Fig. 3a). Interestingly, the ten most frequently used unique CDR3 sequences accounted for 60% of the  $T_H17$  and only 40% of the non- $T_H17$  repertoire (Extended Data Fig. 3b). Furthermore, the dominant CDR3 sequences in individual mice exhibited a clear bias towards either  $T_H17$  or non- $T_H17$  cells (Supplementary Table 1). Many of these CDR3 sequences were shared between mice and were enriched either in  $T_H17$  or in non- $T_H17$  cells in individual mice (Extended Data Fig. 3c).

The finding that intestinal  $T_H17$  cells have a distinct repertoire prompted us to further determine their antigen specificity. Thus, we sorted single T cells from four mice and sequenced their V $\beta$ 14 and paired V $\alpha$  chains (Extended Data Fig. 4a). Notably, each mouse carried some V $\beta$ 14 sequences that were present in several sorted cells, and these sequences strongly biased towards  $T_H17$  or non- $T_H17$  cells (Extended Data Fig. 4b), corroborating our findings from the high-throughput sequencing

analysis. To define the antigen specificity of the TCRs from intestinal  $T_H17$  and non- $T_H17$  cell clones, we expressed a cohort of nineteen pre-dominant clonotypic TCRs (ten  $T_H17$  clones, eight non- $T_H17$  clones, and one neutral clone) in a NFAT-GFP $^+$  hybridoma that can report on TCR signalling<sup>17</sup>. Upon co-culture of the hybridomas with splenic APCs and heat-inactivated mouse intestinal luminal content, several  $T_H17$ -TCR hybridomas, but not non- $T_H17$ -TCR hybridomas, responded to Taconic antigens, and not to Jackson antigens (Extended Data Fig. 4c). Furthermore, when SFB-mono antigens were used, we detected responses from 7/10  $T_H17$ -TCR and the neutral TCR hybridoma, but none of the non- $T_H17$  cell hybridomas (Fig. 2a). These responses were abrogated if the APCs were from MHCII-deficient mice (Extended Data Fig. 4d).

We next sought to identify epitopes recognized by  $T_H17$  cell TCRs using a whole-genome shotgun cloning and expression screen, an unbiased approach previously used to identify T-cell antigens from other bacteria<sup>18</sup> (Extended Data Fig. 5a). One bacterial clone, designated *3F12-E8*, stimulated 7B8 and four other  $T_H17$ -TCR hybridomas (Extended Data Fig. 5b–d). Based on the recent annotation of the SFB genome<sup>19,20</sup>, we assigned the 672-base pair *3F12-E8* insert to an SFB gene (*SFBNYU\_003340*<sup>19</sup>). We confirmed the specificity by cloning the full-length gene and demonstrating that its product stimulated the aforementioned five TCRs, but not any other TCRs (Fig. 2b, left). We further mapped a minimal epitope that stimulated all five TCRs and a shorter 8-amino-acid epitope that stimulated only the 7B8 and 2A6 hybridomas (Extended Data Fig. 5e).

Another expression screen was performed using the 1A7 hybridoma, which along with three other TCRs formed a distinct cluster with an identical V $\alpha$  and highly similar V $\beta$ 14 CDR3 sequences (Extended Data Fig. 6a). A stimulatory clone, designated *2D10-A10* (Extended Data Fig. 6b, c), contained the amino-terminal sequence of another SFB gene (*SFBNYU\_004990*<sup>19</sup>). We mapped the epitope for the 1A7 hybridoma to 9 amino acids (Extended Data Fig. 6d). Both the full-length gene product and a 9-amino-acid peptide stimulated all four TCRs, indicating that

these TCRs indeed recognize the same epitope (Fig. 2b, right). However, the single TCR derived from non- $T_H17$  cells (3F4) displayed a much weaker dose-response to peptide antigen than the other TCRs (Extended Data Fig. 6e).

Thus, eight out of eleven  $V\beta14^+$   $T_H17$ -TCR hybridomas recognized two distinct antigens encoded by SFB (Fig. 2c). Both proteins are unique to SFB, expressed at a medium to high level, and predicted to be secreted or at the cell surface (Fig. 2d). Importantly, primary  $V\beta14^+$   $T_H17$  cells responded to the two immunodominant SFB epitopes (Extended Data Fig. 7a). Although  $V\beta14^+$  cells consistently responded slightly better,  $V\beta14^-$   $T_H17$  cells were also stimulated by SFB (Extended Data Fig. 7b), indicating that these cells respond to other SFB epitopes. An *in silico* search was conducted for potential epitopes within the SFB proteome (Extended Data Fig. 7c, d), which yielded several more stimulatory peptides (Extended Data Fig. 7e). Among these, peptide N5, also derived from *SFB**NYU\_003340*, was a strong stimulator of intestinal  $T_H17$  cells, activating both  $V\beta14^+$  cells and  $V\beta14^-$  cells (Extended Data Fig. 7f). Thus, in the small intestine, SFB is the dominant antigen source for polyclonal  $T_H17$  cells, but for few, if any, non- $T_H17$  cells.

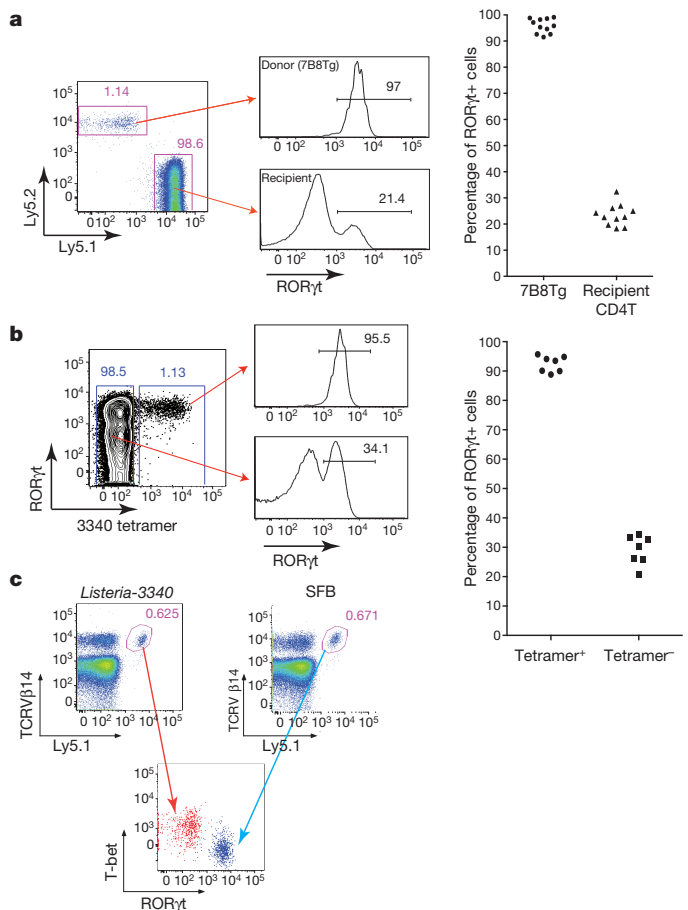
We then asked what happens to T cells expressing SFB-specific TCRs. We generated 7B8 TCR transgenic mice (7B8Tg)<sup>21</sup>, and transferred naive T cells from these mice into isotype-congenic B6 mice<sup>22</sup>. After one week, we readily detected donor-derived T cells in the SILP of mice that had been exposed to SFB, whereas they were completely absent in SFB-deficient recipients (Extended Data Fig. 8a). Remarkably, almost all donor-derived cells became positive for ROR $\gamma$ t (Fig. 3a). Similar results were obtained upon transfer of T cells from two other TCR transgenic strains (1A2Tg and 5A11Tg) into SFB-colonized recipient mice (Extended Data Fig. 8b). The donor-derived T cells lacked expression of the transcription factors associated with alternative CD4 T-cell programs (for example, FOXP3, GATA3, and T-bet) (Extended Data Fig. 8c).

To visualize endogenous SFB-antigen-specific T cells, we produced MHCII-tetramers containing peptide A6 from *SFB**NYU\_003340* (3340-A6 tetramer)<sup>23</sup>. The I-A<sup>b</sup>/3340-A6 tetramer specifically stained GFP<sup>+</sup> SILP CD4<sup>+</sup> T cells from SFB-colonized *Il23r*<sup>GFP/+</sup> mice (Extended Data Fig. 8d). Furthermore, a sizable population of I-A<sup>b</sup>/3340-A6 tetramer-positive cells was present in B6 Taconic, but not in B6 Jackson mice (Extended Data Fig. 8e), and these cells were uniformly ROR $\gamma$ t-positive, indicating that they were SFB-elicited  $T_H17$  cells (Fig. 3b).

We next aimed to determine whether polarization of the antigen-specific  $T_H17$  cells in response to SFB colonization is dictated by the nature of the antigenic protein or properties of the microbe. *Listeria monocytogenes*, an enteric pathogenic bacterium that also colonizes the small intestine, typically elicits a  $T_H1$  response<sup>24</sup>. Mice were orally infected with *L. monocytogenes* expressing *SFB**NYU-003340* (*Listeria*-3340) (Extended Data Fig. 8f) or SFB before intravenous transfer of 7B8Tg T cells. 7B8Tg T cells accumulated in the SILP of both sets of mice, but, importantly, they expressed T-bet rather than ROR $\gamma$ t when the hosts were colonized with *Listeria*-3340 (Fig. 3c).

To further investigate a relationship between the fate of SILP T helper cells and the bacterial origins of antigens, we transferred 7B8Tg T cells into mice that were colonized with both SFB and *Listeria* and simultaneously tracked CD4<sup>+</sup> T-cell responses specific for both bacteria in the SILP using the Ly5.1<sup>+</sup> congenic marker for 7B8Tg cells and listeriolysin O (LLO)-tetramers that stain endogenous *Listeria*-specific T cells derived from the host (Extended Data Fig. 9a). In the presence of both  $T_H17$ - and  $T_H1$ -inducing bacteria, 7B8Tg T cells expressed ROR $\gamma$ t, but not T-bet, whereas LLO-tetramer<sup>+</sup> cells expressed T-bet, but not ROR $\gamma$ t (Fig. 4a and Extended Data Fig. 9b, c). This result is in contrast to the  $T_H1$  polarization of TCR transgenic T cells specific for the commensal CBir1 flagellin antigen observed upon infection with the protozoan parasite *Toxoplasma gondii*<sup>25</sup>, a  $T_H1$ -inducing intestinal pathogen. This suggests that, unlike CBir1-encoding *Clostridia*, SFB has the ability to direct a dominant signal specialized for induction of  $T_H17$  cells.

SFB colonization of the small intestine is potentially beneficial, attenuating pathogenic bacteria-induced colitis<sup>8</sup>, but it can also trigger or

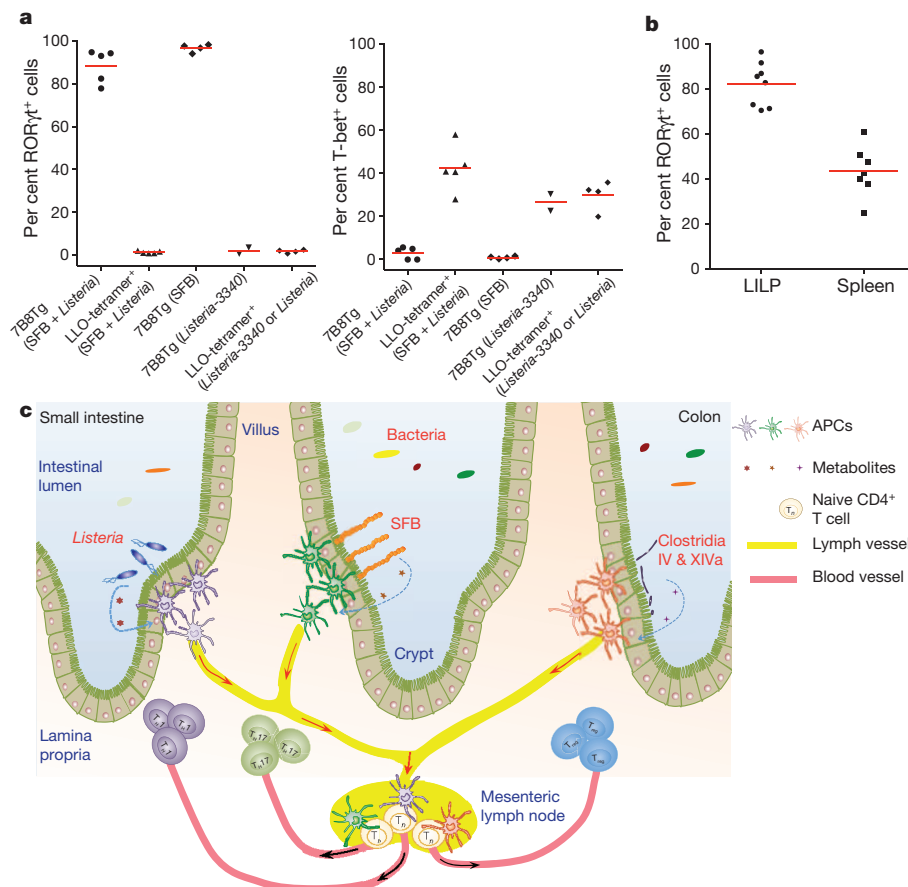


**Figure 3 | SFB-specific T cells become  $T_H17$  cells in the SILP.** **a**, 7B8Tg cells (Ly5.2) were transferred into SFB-colonized mice (Ly5.1), and SILP T cells were analysed after 8–15 days. Left, representative FACS plots. Right, analysis of multiple animals (one symbol per animal). **b**, I-A<sup>b</sup>/3340-A6 tetramer stain of SILP T cells from SFB-colonized B6 mice. Left, representative FACS plots. Right, analysis of multiple animals (one symbol per animal). **c**, 7B8Tg cells (Ly5.1) were transferred into Ly5.2 congenic hosts orally colonized with *Listeria*-3340 or SFB. Seven days after transfer, donor-derived cells in the SILP were analysed. Analyses of cells from different mice are shown in the composite FACS plot. The results are representative of three experiments.

exacerbate systemic autoimmune disease<sup>10,11</sup>, raising the question as to whether SFB-specific  $T_H17$  cells can circulate beyond the small intestine. We examined the colons and spleens of SFB-positive recipients of 7B8Tg naive T cells, and found these cells in both organs. Importantly, more than 80% of these SFB-specific T cells in colon and 40% in spleen expressed ROR $\gamma$ t (Fig. 4b). Consistent with this result, staining of endogenous T cells from Taconic mice revealed 3340-A6 tetramer-positive cells in the large intestine and most of these cells expressed ROR $\gamma$ t (Extended Data Fig. 10a, b).

Our results therefore indicate that intestinal antigen-specific CD4<sup>+</sup> T cells differentiate to become either  $T_H1$  or  $T_H17$  cells, depending on which luminal bacterium delivers the antigen. We propose a deterministic model for T helper cell differentiation whereby the bacterial context of cognate antigen delivery dictates the fate of the antigen-specific T cells (Fig. 4c). Our work opens the way towards elucidating the mechanisms of  $T_H17$  cell induction by microbiota and of how gut-induced  $T_H17$  cells can contribute to distal organ-specific autoimmune disease. In addition, it serves as a guide for future studies of human commensal-specific pro-inflammatory T cells that are thought to contribute to autoimmune diseases such as rheumatoid arthritis<sup>26</sup>. Finally, the demonstration of controlled polarized T-cell responses towards commensal bacteria offers the potential for novel approaches towards mucosal vaccination.





**Figure 4 | TCR specificity for distinct luminal bacteria underlies divergent T helper cell differentiation in the SILP.** **a**, T<sub>H</sub>17 (RORγt) versus T<sub>H</sub>1 (T-bet) differentiation of SFB- (7B8Tg) and *Listeria* (LLO-tetramer)-specific CD4<sup>+</sup> T cells in mice colonized with either or both bacteria. Each symbol represents cells from one animal. **b**, Proportions of donor-derived 7B8Tg T cells that express RORγt in the colon and spleen of SFB-colonized mice. LILP, large intestine lamina propria. **c**, Model for intestinal niches that promote diverse microbiota-dependent CD4<sup>+</sup> effector T-cell programs. Microbial signals may induce polarizing cytokines or preformed niche-specific antigen-presenting cells may interact with different T-cell-inducing bacteria.

## METHODS SUMMARY

**Mice.** All mice were housed in the animal facility of The Skirball Institute of Bio-molecular Medicine at the New York University School of Medicine. Experimental protocols were approved by the Institutional Animal Care and Use Committee. C57BL/6 mice were purchased from Taconic Farm (B6 Taconic) or the Jackson Laboratory (B6 Jackson). *Il23r<sup>GFP</sup>* mice<sup>16</sup>, a gift from M. Oukka, were maintained by breeding with B6 Taconic mice. 7B8Tg, 1A2Tg and 5A11Tg SFB-specific TCR transgenic (Tg) mice were generated as previously described<sup>21</sup> and kept with SFB-minus flora. For adoptive transfer, naive Tg T cells (CD62L<sup>hi</sup> CD44<sup>lo</sup> Vβ14<sup>+</sup> CD4<sup>+</sup> CD3<sup>+</sup>) were sorted from the spleen and were injected intravenously into congenic recipient mice<sup>22</sup>.

**Generation of TCR hybridomas.** Retroviruses carrying an expression cassette encoding TCRα, TCRβ, and CD4 were used to infect the NFAT-GFP 58α<sup>−</sup>β<sup>−</sup> hybridoma cell line<sup>17</sup>.

**Construction and screen of whole-genome shotgun library of SFB.** The shotgun library was prepared with a procedure modified from a previous study<sup>18</sup>. The library is estimated to contain 10<sup>4</sup> clones. The expression of exogenous proteins was induced by isopropylthiogalactoside for 4 h. For antigen screening, pools of heat-killed bacteria (~30 clones per pool) were added to a co-culture of APCs and hybridomas.

**MHCII tetramer production and staining.** MHCII/3340-A6 tetramer was produced as previously described<sup>23</sup>. SILP T cells were incubated at room temperature for 60 min with fluorochrome-labelled tetramer (10 nM) before staining with relevant antibodies at 4 °C.

**Heterologous expression of SFBNYU\_003340 in *Listeria monocytogenes*.** The entire coding region of SFBNYU\_003340, including its predicted signal sequence, was sub-cloned into the *Listeria* expression vector pIMK2<sup>27</sup>. The resultant plasmid was transformed into electrocompetent *Listeria monocytogenes* strain 10403S-*inla*<sup>tm</sup> and plated on selective medium containing kanamycin (50 μg ml<sup>−1</sup>)<sup>28</sup>.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 December 2013; accepted 25 March 2014.

Published online 13 April 2014.

1. Bettelli, E., Korn, T., Oukka, M. & Kuchroo, V. K. Induction and effector functions of T<sub>H</sub>17 cells. *Nature* **453**, 1051–1057 (2008).

2. McGeachy, M. J. & Cua, D. J. The link between IL-23 and Th17 cell-mediated immune pathologies. *Semin. Immunol.* **19**, 372–376 (2007).
3. Littman, D. R. & Rudensky, A. Y. Th17 and regulatory T cells in mediating and restraining inflammation. *Cell* **140**, 845–858 (2010).
4. Ivanov, I. I. *et al.* The orphan nuclear receptor RORγt directs the differentiation program of proinflammatory IL-17<sup>+</sup> T helper cells. *Cell* **126**, 1121–1133 (2006).
5. Ivanov, I. I. *et al.* Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. *Cell Host Microbe* **4**, 337–349 (2008).
6. Atarashi, K. *et al.* ATP drives lamina propria T<sub>H</sub>17 cell differentiation. *Nature* **455**, 808–812 (2008).
7. Atarashi, K. *et al.* T<sub>reg</sub> induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500**, 232–236 (2013).
8. Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).
9. Gaboriau-Routhiau, V. *et al.* The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* **31**, 677–689 (2009).
10. Wu, H. J. *et al.* Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. *Immunity* **32**, 815–827 (2010).
11. Lee, Y. K., Menezes, J. S., Umesaki, Y. & Mazmanian, S. K. Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc. Natl Acad. Sci. USA* **108** (Suppl 1), 4615–4622 (2011).
12. Ivanov, I. I. & Honda, K. Intestinal commensal microbes as immune modulators. *Cell Host Microbe* **12**, 496–508 (2012).
13. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012).
14. Schnupf, P., Gaboriau-Routhiau, V. & Cerf-Bensussan, N. Host interactions with Segmented Filamentous Bacteria: An unusual trade-off that drives the post-natal maturation of the gut immune system. *Semin. Immunol.* **25**, 342–351 (2013).
15. Lochner, M. *et al.* Restricted microbiota and absence of cognate TCR antigen leads to an unbalanced generation of Th17 cells. *J. Immunol.* **186**, 1531–1537 (2011).
16. Awasthi, A. *et al.* Cutting edge: IL-23 receptor GFP reporter mice reveal distinct populations of IL-17-producing cells. *J. Immunol.* **182**, 5904–5908 (2009).
17. Ise, W. *et al.* CTLA-4 suppresses the pathogenicity of self antigen-specific T cells by cell-intrinsic and cell-extrinsic mechanisms. *Nature Immunol.* **11**, 129–135 (2010).

18. Sanderson, S., Campbell, D. J. & Shastri, N. Identification of a CD4<sup>+</sup> T cell-stimulating antigen of pathogenic bacteria by expression cloning. *J. Exp. Med.* **182**, 1751–1757 (1995).
19. Sczesnak, A. *et al.* The genome of Th17 cell-inducing segmented filamentous bacteria reveals extensive auxotrophy and adaptations to the intestinal environment. *Cell Host Microbe* **10**, 260–272 (2011).
20. Prakash, T. *et al.* Complete genome sequences of rat and mouse segmented filamentous bacteria, a potent inducer of Th17 cell differentiation. *Cell Host Microbe* **10**, 273–284 (2011).
21. Kouskoff, V., Signorelli, K., Benoist, C. & Mathis, D. Cassette vectors directing expression of T cell receptor genes in transgenic mice. *J. Immunol. Methods* **180**, 273–280 (1995).
22. Kearney, E. R., Pape, K. A., Loh, D. Y. & Jenkins, M. K. Visualization of peptide-specific T cell immunity and peripheral tolerance induction *in vivo*. *Immunity* **1**, 327–339 (1994).
23. Moon, J. J. *et al.* Naive CD4<sup>+</sup> T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* **27**, 203–213 (2007).
24. Hsieh, C. S. *et al.* Development of TH1 CD4<sup>+</sup> T cells through IL-12 produced by *Listeria*-induced macrophages. *Science* **260**, 547–549 (1993).
25. Hand, T. W. *et al.* Acute gastrointestinal infection induces long-lived microbiota-specific T cell responses. *Science* **337**, 1553–1556 (2012).
26. Scher, J. U. *et al.* Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* **2**, e01202 (2013).
27. Monk, I. R., Gahan, C. G. & Hill, C. Tools for functional postgenomic analysis of *Listeria monocytogenes*. *Appl. Environ. Microbiol.* **74**, 3921–3934 (2008).
28. Xayarath, B., Marquis, H., Port, G. C. & Freitag, N. E. *Listeria monocytogenes* CtaP is a multifunctional cysteine transport-associated protein required for bacterial pathogenesis. *Mol. Microbiol.* **74**, 956–973 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank S. Yong Kim for generating TCR transgenic mice, A. Viale for 454 pyrosequencing, R. Myers for RNA-seq, N. Freitag for providing the *Listeria* strain and expression vector, Y. Umesaki for SFB samples, and K. Murphy for providing the 58 $\alpha$   $\beta$  hybridoma line. The Immune Monitoring Core New York University (NYU) is supported in part by grant UL1 TR00038 from the National Center for Advancing Translational Sciences and grant 5P30CA016087-32 from the National Cancer Institute; the NYU Histology Core is supported in part by grant 5P30CA016087-32 from the National Cancer Institute. M.K.J. was supported by grant R01 AI039614 from the NIH. Y.Y. was supported by the Arthritis National Research Foundation. M.X. is supported by the Irvington Institute fellowship program of the Cancer Research Institute. D.R.L. is a Howard Hughes Medical Institute Investigator.

**Author Contributions** Y.Y. and D.R.L. designed the experiments and wrote the manuscript with input from the co-authors. Y.Y., M.B.T., M.X., C.N., A.C., X.L. and J.-J.L. performed most analyses. M.B.T. constructed TCR hybridomas. M.X. developed SFB-specific antibodies. M.G., H.X. and J.J.L. did TCR pyrosequencing analysis. J.L.L. and M.K.J. developed tetramers. F.A. and V.J.T. generated transgenic *Listeria*. A.S. performed RNA-seq analysis of SFB.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.R.L. ([dan.littman@med.nyu.edu](mailto:dan.littman@med.nyu.edu)).

## METHODS

**Mice.** C57BL/6 mice were purchased from Taconic Farm (B6 Taconic) or Jackson Laboratory (B6 Jackson). *Il23r<sup>GFP</sup>* mice<sup>16</sup> were provided by M. Oukka and maintained by breeding with B6 Taconic mice. Ly5.1 mice (*B6.SJL-Ptprca Pepcb/BoyJ*) and MHCII-deficient mice (*B6.129S2-H2<sup>dAb1-Ea</sup>/J*) were from Jackson Laboratory.

**Antibodies and flow cytometry.** The following antibodies were from eBiosciences, BD Pharmingen or BioLegend: Vβ2 (B20.6), Vβ3 (KJ25), Vβ4 (KT4), Vβ5 (MR9-4), Vβ6 (RR4-7), Vβ7 (TR310), Vβ8 (F23.1), Vβ8.1/8.2 (MR5-2), Vβ8.3 (8C1), Vβ10 (B21.5), Vβ11 (CTVB11), Vβ12 (MR11-1), Vβ14 (14-2), CD3 (145-2C11), CD4 (RM4-5), CD25 (PC61), Ly5.1 (A20), Ly5.2 (104), MHCII (M5/114), RORγt (AFKJS-9 or B2D), FOXP3 (FJK-16s), T-bet (eBio4B10), GATA3 (TWAJ), IL-17A (eBio17B7) and IFN-γ (XM61.2). Flow cytometric analysis was performed on an LSR II (BD Biosciences) or an Aria II (BD Biosciences) and analysed using FlowJo software (Tree Star). DAPI (Sigma) was used to exclude dead cells.

**T-cell preparation and staining.** Small intestine lamina propria were minced and then incubated for 30 min at 37 °C with collagenase D (1 mg ml<sup>-1</sup>; Roche), dispase (0.05 U ml<sup>-1</sup>; Worthington) and DNase I (100 µg ml<sup>-1</sup>; Sigma). Lymphocytes were collected at the interface of a 40%/80% Percoll gradient (GE Healthcare). Cells were stained for surface markers, followed by fixation and permeabilization (eBioscience).

**Calculating enrichment scores.** An enrichment score for a given Vβ in IL-23R (GFP)<sup>+</sup> cells is defined as the equation of (per cent of Vβ<sup>+</sup> cells in the GFP-positive fraction)/(per cent of Vβ<sup>+</sup> cells in the GFP-negative fraction). For example, for Extended Fig. 2b, Vβ14 enrichment was calculated as (7.45/(7.45 + 26.2))/(4.48/(4.48 + 61.8)) or 3.3. A score > 1 means a positive enrichment and a score ≈ 1 means no enrichment.

**High-throughput TCR sequencing.** The SILP cells from *Il23r<sup>GFP/+</sup>* mice were stained for surface markers and Vβ14<sup>+</sup> CD4<sup>+</sup> T cells were sorted on the Aria II. For each sample, we collected about 2 × 10<sup>4</sup> cells (2.17 ± 0.43 × 10<sup>4</sup> cells for GFP<sup>+</sup> T<sub>H</sub>17 cells and 2.38 ± 0.54 × 10<sup>4</sup> cells for GFP<sup>-</sup> non-T<sub>H</sub>17 cells). Cells were lysed in TRIzol reagent (Invitrogen) and RNA was extracted following the manufacturer's instruction. RNA precipitation was aided with GlycoBlue (Invitrogen). Complementary DNAs were prepared with a reverse transcription kit (USB). Vβ14 PCRs were performed using barcoded oligonucleotides. PCR products from 16 samples were quantified on NanoDrop. Equal amounts of barcoded PCR product were mixed and sequenced using a 454 GS Junior system (Roche). The raw sequencing data was first aligned using the high-throughput analysis tool provided by IMGT<sup>29</sup>. We obtained 6,647 ± 954 reads for T<sub>H</sub>17 cells and 5,573 ± 889 reads for non-T<sub>H</sub>17 cells. CDR3 usage was further computed with Perl-based scripts developed in-house. The T<sub>H</sub>17 samples had 340–772 unique Vβ14 CDR3 sequences and the non-T<sub>H</sub>17 samples had 849–2,148 unique Vβ14 CDR3 sequences.

**Single-cell TCR sequencing.** The SILP cells from *Il23r<sup>GFP/+</sup>* mice were stained for surface markers. GFP<sup>+</sup> and GFP<sup>-</sup> Vβ14<sup>+</sup> CD4<sup>+</sup> T cells were sorted on the BD Aria II and deposited at one cell per well into 96-well PCR plates preloaded with 5 µl reverse transcription mix (USB). Immediately after sorting, whole plates were incubated at 50 °C for 60 min for cDNA synthesis. Half of the cDNA was used for Vβ14 PCR using forward primer 5'-ACGACCAATTCATCCTAAGCAC-3' and reverse primer 5'-AAGCACACGAGGTAGCCT-3'. To retrieve Vα sequences, the other half of cDNA was pre-amplified for 16 cycles using a mix of twenty-one forward primers<sup>30</sup> (each modified by adding a 5' extended anchor sequence: TAATACGACTCACTATAGGG) and a reverse primer 5'-CATGTCCAGCACA GTTTTGTCAGT-3'. The primary Vα PCR products were diluted and subjected to a second round PCR using forward primer 5'-TAATACGACTCACTATAG GG-3' and reverse primer 5'-GTCAAAGTCGGTGAACAGGC-3'. PCRs were performed in a LightCycler 480 (Roche). PCR products were cleaned up with ExoSap-IT reagent (USB) and Sanger sequencing was performed by Macrogen. In nearly all cases, for cells with the same Vβ14 sequence, we retrieved a single unique Vα sequence, indicating that these cells were clonotypically identical.

**Generation of TCR hybridomas.** The NFAT-GFP 58α<sup>-</sup>β<sup>-</sup> hybridoma cell line<sup>17</sup> was provided by K. Murphy. To reconstitute TCRs, we used a self-cleavage sequence of 2A to link cDNAs of TCRα and TCRβ generated from annealing of overlapping oligonucleotides (TCRα-p2A-TCRβ) and shuttled the cassette into a modified MigR1 retrovector in which IRES-GFP was replaced with IRES-mCD4. Then retroviral vectors were transfected into Phoenix E packaging cells using Lipofectamine 2000 (Invitrogen). Hybridoma cells were transduced with viral supernatants in the presence of polybrene (8 µg ml<sup>-1</sup>) by spin infection for 90 min at 32 °C. Transduction efficiencies were monitored by checking mCD3 surface expression on day 2. We generated nineteen hybridomas for predominant clonotypic TCRs whose Vα and Vβ sequences were retrieved from single-cell TCR sequencing (no. of Vβ ≥ 6 for T<sub>H</sub>17 biased clones, and no. of Vβ ≥ 4 for non-T<sub>H</sub>17 biased clones, no. of Vα ≥ 3. Note that a single unique Vα was identified for every Vβ). We also generated the OTII hybridoma using TCR sequences provided by F. Carbone (chicken ovalbumin antigen-specific, I-A<sup>b</sup>-restricted).

**Assay for hybridoma activation.** To prepare antigen-presenting cells, splenocytes from B6 mice that were injected intraperitoneally with 8 × 10<sup>6</sup> FLT3-B16 melanoma cells 10 days before were positively enriched for CD11c<sup>+</sup> cells using MACS LS columns (Miltenyi). 10<sup>4</sup> hybridoma cells were incubated with 2 × 10<sup>5</sup> APCs and autoclaved antigens from intestinal luminal contents or faecal material for two days. GFP induction in the hybridomas (CD3<sup>+</sup> fraction) was analysed by flow cytometry.

**Construction and screen of whole-genome shotgun library of SFB.** The shotgun library was prepared with a procedure modified from a previous study<sup>18</sup>. In brief, genomic DNA was purified from the feces of SFB-monoassociated mice by phenol:chloroform extraction. DNA was subjected to whole-genome amplification with the REPLI-g kit (Qiagen) following the manufacturer's instructions. Amplified materials were partially digested with Sau3A (NEB), then ligated with the BamHI-linearized pGEX-4T1 expression vector (GE Healthcare). Ligation products were introduced into competent Stb3 cells (Invitrogen). To ensure the quality of the library, we sequenced the inserts of randomly picked colonies. All the sequences were mapped to the SFB genome. The library is estimated to contain 10<sup>4</sup> clones. We grew bacteria in 96-well deepwell plates (VWR) with AirPort micro-porous cover (Qiagen). The expression of exogenous proteins was induced by isopropylthiogalactoside for 4 h. Then bacteria were heat-killed by incubating at 70 °C for 1 h, and stored at -20 °C until use. For antigen screens, pools of bacterial clones (~30 clones per pool) were added to a co-culture of APCs and hybridomas. Clones within the positive pools were screened individually against the hybridoma bait. Finally, the inserts of positive clones were subjected to Sanger sequencing. The sequences were blasted against the SFB genome and aligned to annotated open reading frames.

**Epitope mapping.** We expressed overlapping fragments spanning the active ORF using the pGEX-4T1 bacterial expression system, and colonies were used to stimulate the relevant hybridoma. This process was repeated until we identified minimal fragments conferring antigenicity. The mapping was further verified by stimulating hybridomas with synthetic peptides (Genescript).

**RNA-seq analysis of the SFB transcriptome.** Wild-type B6 mice from Jackson Laboratory or Taconic Farm, confirmed for the presence or absence of SFB by quantitative PCR<sup>19</sup>, were used for microbiome transcriptome analysis. Within five minutes after sacrifice, the terminal ileum of each mouse was resected and luminal contents were squeezed with sterile forceps into a mortar cooled with liquid nitrogen. 1 ml nuclease-free TE was washed through the ileum into the mortar. The total luminal contents and washing were then ground to a fine powder with a pestle cooled with liquid nitrogen and kept on dry ice. The powder was then transferred to 15 ml TRIzol (Life Technologies) in a 50 ml falcon tube and vortexed. The manufacturer's protocol was then used to obtain RNA. The resulting RNA was extracted twice with acid phenol-chloroform, precipitated, treated with Ambion Turbo DNA-free, and cleaned-up with an RNeasy column to yield RNA with an undetectable concentration of DNA by Qubit. A portion of this RNA was treated once with Epicentre's Ribo-Zero rRNA removal kit, using equal volumes of specific oligonucleotides from the Meta-bacteria and Human/mouse/rat kits. An Illumina RNA-seq library was prepared from these samples using a previously-described strand-specific Nextera protocol. The resulting reads were aligned to the SFBNYU genome with Bowtie<sup>31</sup> and transcript abundance was estimated using Cufflinks<sup>32</sup> with default parameters.

**Production of anti-SFB antibody and immunostaining.** The cDNA fragments corresponding to amino acids 43–359 (3340N) and 734–1060 (3340C) of *SFBNYU\_003340* were cloned into the pGEX6p1 expression vector. Recombinant proteins fused to N-terminal glutathione-S-transferase (GST) were expressed in *E. coli* BL21, purified with glutathione Sepharose 4B (GE), and were released with PreScission protease (GE). The flow-through fractions containing polypeptides without the GST tag were collected as immunogen. Rabbit polyclonal antibodies against both polypeptides were raised by Covance. For immunostaining, bacteria were fixed with 2% paraformaldehyde, followed by washing with 0.5% TritonX-100. Bacteria were incubated sequentially with primary antibody (1:1 mix of the two rabbit anti-3340 antibodies) and phycoerythrin-conjugated goat anti-rabbit antibody.

**Activation of polyclonal SILP T<sub>H</sub>17 cells.** GFP<sup>+</sup> and GFP<sup>-</sup> SILP CD4<sup>+</sup> T cells sorted from *Il23r<sup>GFP/+</sup>* mice were incubated with 2 × 10<sup>5</sup> APCs (CD11c<sup>+</sup> cells purified from the spleen) and indicated stimuli in complete RPMI medium supplemented with IL-2 (10 U ml<sup>-1</sup>) and IL-7 (5 ng ml<sup>-1</sup>) for 2–3 days. Cells were collected and stained with Vβ-specific antibodies. Forward scatter increment, as readout for cell activation, was analysed by FACS.

**IL-17A ELISPOT assay.** IL-17A ELISPOT was performed with a mouse/rat IL-17A ELISPOT Ready-SET-Go! kit (eBioscience). Dots were automatically enumerated with ImmunoSpot software (Version 5.0).

**MHCII tetramer production and staining.** I-A<sup>b</sup>/3340-A6 tetramer was produced as previously described<sup>23</sup>. Briefly, QFSGAVPNKTD, an immunodominant epitope from *SFBNYU\_0033400*, covalently linked to I-A<sup>b</sup> via a flexible linker, was produced in *Drosophila* S2 cells. Soluble pMHCII monomers were purified, biotinylated,



and tetramerized with phycoerythrin- or allophycocyanin-labelled streptavidin. To stain endogenous cells, SILP cells were first resuspended in FACS buffer with FcR block, 2% mouse serum and 2% rat serum. Then tetramer was added (10 nM) and incubated at room temperature for 60 min. Cells were washed and followed by regular staining at 4 °C. I-A<sup>b</sup>/2W and I-A<sup>b</sup>/LLO tetramers were previously described<sup>23,33</sup>.

**Generation of T<sub>H</sub>17-TCRTg mice.** TCR sequences of 7B8, 1A2 and 5A11 were cloned into the pT $\alpha$  and pT $\beta$  vectors kindly provided by D. Mathis<sup>21</sup>. TCR transgenic animals were generated by the Rodent Genetic Engineering Core at the New York University School of Medicine. Positive pups were genotyped by PCR and kept on SFB-minus flora.

**Adoptive transfer.** Spleens from 7B8Tg mice were collected and dissociated. Red blood cells were lysed using ACK lysis buffer (Lonza). Naive Tg T cells (CD62L<sup>hi</sup> CD44<sup>lo</sup> V $\beta$ 14<sup>+</sup> CD4<sup>+</sup> CD3<sup>+</sup>) were sorted on a BD Aria II. Cells were transferred into congenic Ly5.1 recipient mice by retro-orbital injection. In some experiments, we used Ly5.1/Ly5.2 TCRTg mice as donor and transferred naive Tg T cells to congenic Ly5.2 recipient mice.

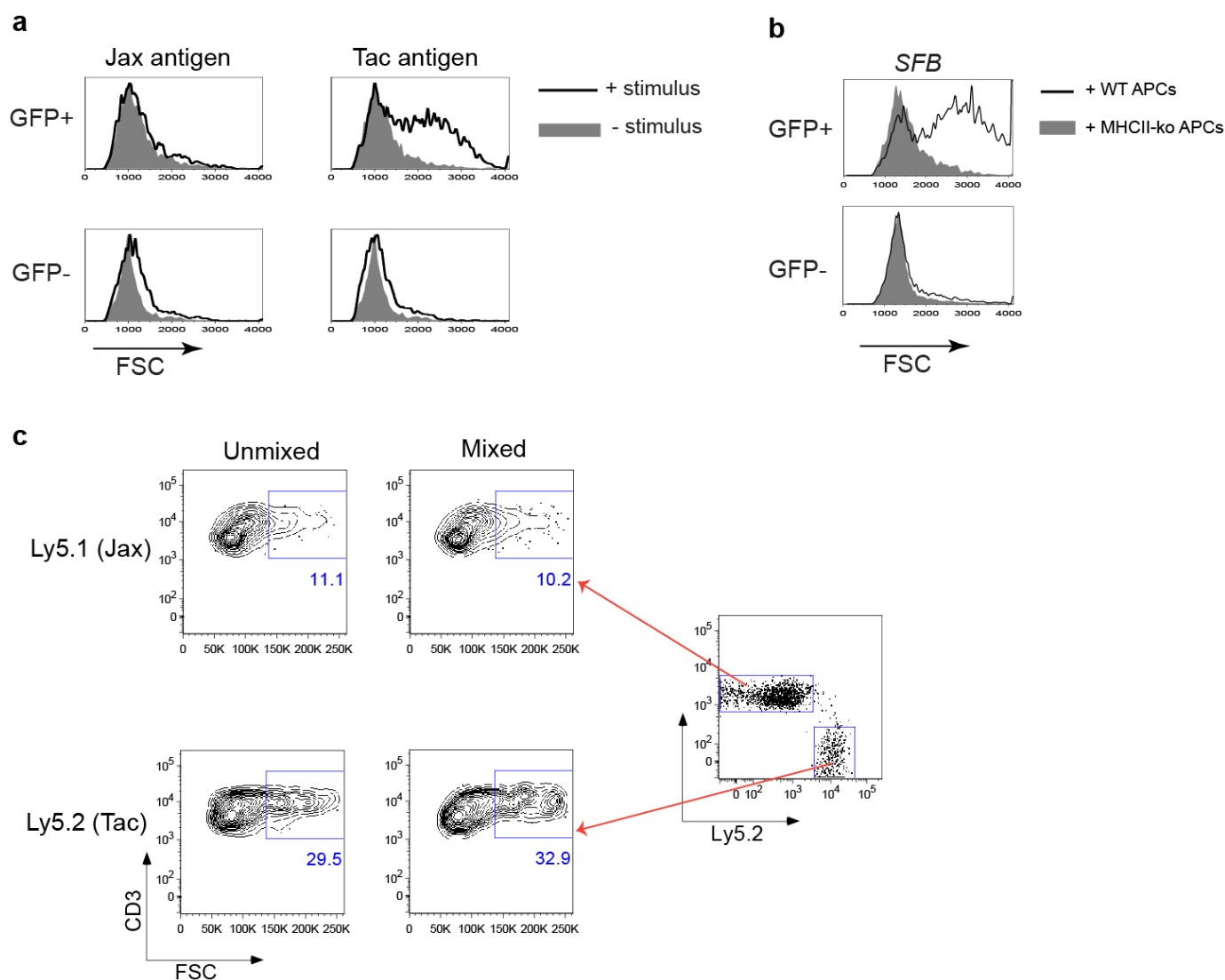
**Heterologous expression of SFBNYU\_003340 in *Listeria monocytogenes*.** To generate strains of *L. monocytogenes* that express the SFBNYU\_003340 antigen, the entire coding region including its predicted signal sequence was PCR-amplified from a plasmid containing the SFBNYU\_003340 gene. The resultant PCR product was digested and sub-cloned into the *Listeria* expression vector pIMK2 (provided by C. Hill), allowing the gene to be expressed under the synthetic promoter P<sub>help</sub> (High expression promoter in *L. monocytogenes*)<sup>27</sup>. The resultant plasmid designated pIMK2-3340 was transformed into electrocompetent *Listeria monocytogenes* strain 10403S-*inlA*<sup>m</sup> (provided by N. E. Freitag) and plated on selective medium containing kanamycin (50  $\mu$ g ml<sup>-1</sup>)<sup>28</sup>. pIMK2 is a derivative of the plasmid pPL2 and stably integrates in single copy within the tRNA<sup>Arg</sup> gene following electroporation<sup>34</sup>. The integrity of the SFBNYU\_003340 gene was validated by PCR and expression confirmed by Coomassie staining of *L. monocytogenes* exoproteins.

**Oral infection with SFB and *L. monocytogenes*.** For SFB colonization, we dissolved in sterile PBS fresh faecal pellets collected from *Il23r<sup>GFP/GFP</sup> Rag2<sup>-/-</sup>* mice that have highly elevated levels of SFB, and infected mice by oral gavage. For *L. monocytogenes* colonization, we grew *Listeria*-3340 and *Listeria*-empty in brain heart infusion medium and infected mice orally with 10<sup>9</sup> colony forming units.

**Bioinformatic analysis.** Protein predictions were made by bioinformatic tools, including Psort (Version 3.0)<sup>35</sup> and Cello (Version 2.5)<sup>36</sup> for localization prediction, and IEDB (Immune Epitope Database) for MHCII binding affinity prediction.

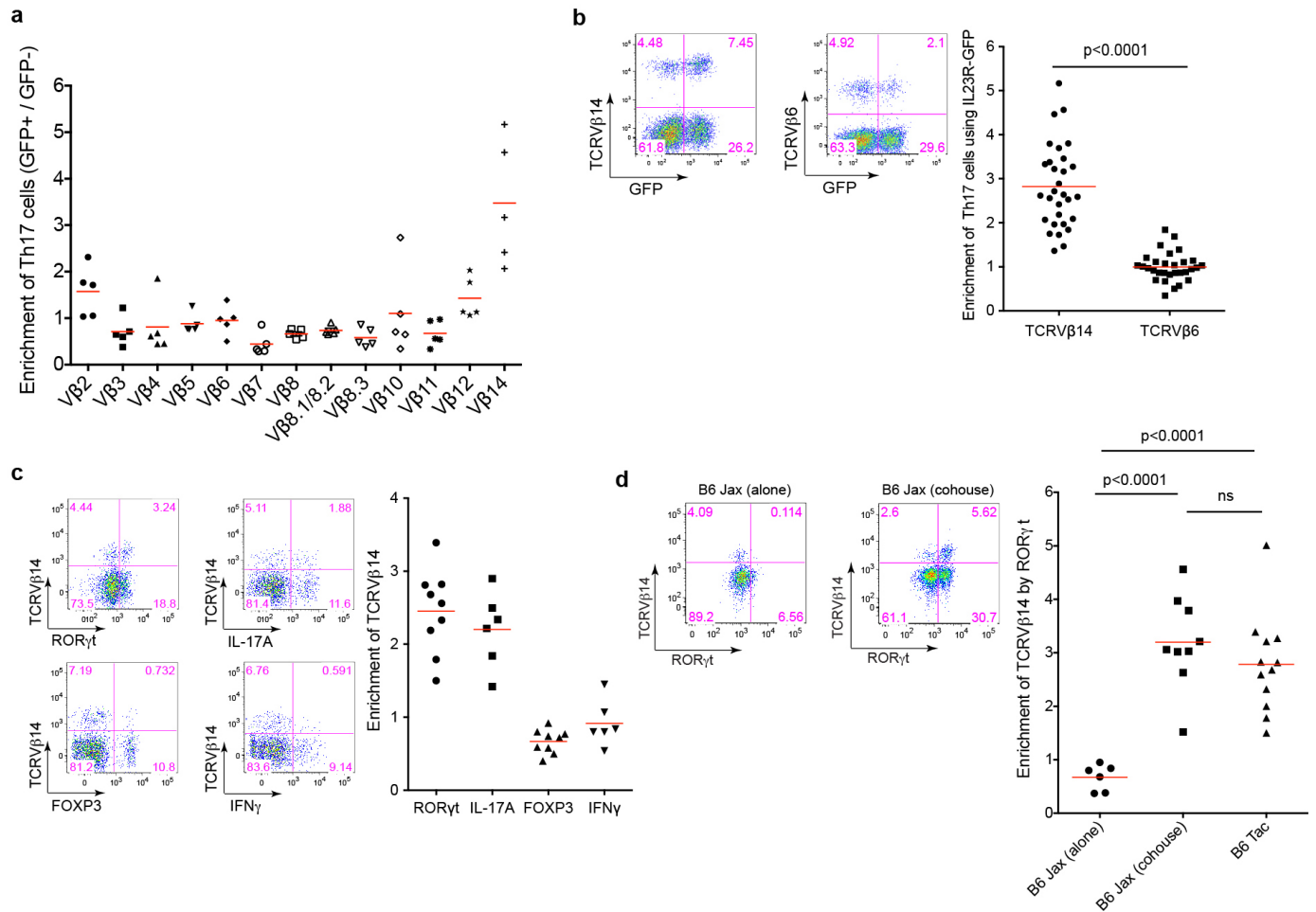
**Statistical analysis.** All analyses were performed using GraphPad Prism (Version 6.0). Differences were considered to be significant at *P* values <0.05.

29. Alamyar, E., Giudicelli, V., Li, S., Duroux, P. & Lefranc, M. P. IMGT/HighV-QUEST: the IMGT(R) web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.* **8**, 26 (2012).
30. Currier, J. R. & Robinson, M. A. Spectratype/immunoscope analysis of the expressed TCR repertoire. *Curr. Protocols Immunol.* **Chapter 10**, Unit 10.28 (2001).
31. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
32. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325–2329 (2011).
33. Tubo, N. J. *et al.* Single naive CD4<sup>+</sup> T cells from a diverse repertoire produce different effector cell types during infection. *Cell* **153**, 785–796 (2013).
34. Lauer, P., Chow, M. Y., Loessner, M. J., Portnoy, D. A. & Calendar, R. Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. *J. Bacteriol.* **184**, 4177–4186 (2002).
35. Yu, N. Y. *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615 (2010).
36. Yu, C. S., Chen, Y. C., Lu, C. H. & Hwang, J. K. Prediction of protein subcellular localization. *Proteins* **64**, 643–651 (2006).



**Extended Data Figure 1 | Stimulation of SILP  $T_H17$  cells requires intestinal microbiota antigen presentation.** **a**, Intestinal  $GFP^+$   $CD4^+$   $T$  cells from  $Il23^{GFP/+}$  mice stimulated with faecal material from Jackson and Taconic mice in the presence of syngeneic splenic APCs. Forward scatter was evaluated after 2 days. **b**,  $T_H17$  cell activation by faecal material from SFB-monoassociated

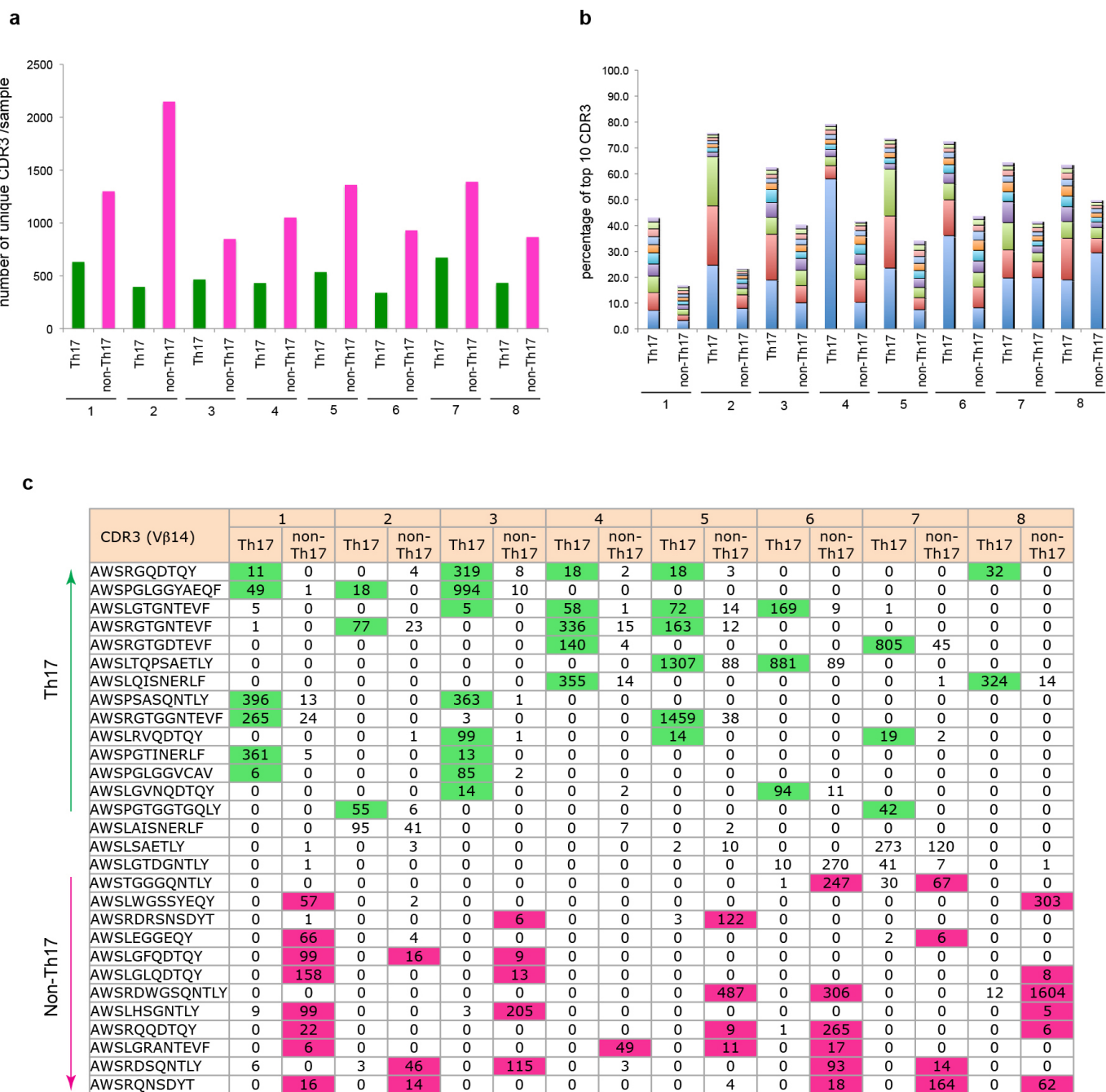
mice in the presence of APCs sufficient (WT) or deficient (KO) for MHC class II. **c**, Evaluation of potential activation of bystander  $CD4^+$   $T$  cells upon stimulation with SFB antigen. SILP  $CD4^+$   $T$  cells from mice with Jackson flora (Ly5.1) and Taconic flora (Ly5.2) were co-cultured or stimulated separately with APCs and SFB-monoassociated faecal material, and FSC was evaluated.



**Extended Data Figure 2 | Microbiota-dependent TCR usage bias among SILP TH17 cells.** **a**, SILP CD4<sup>+</sup> T cells from *Il23r*<sup>GFP/+</sup> mice were analysed for utilization of Vβs in TH17 cells versus non-TH17 cells. Ratios of the percentage of each TCR Vβ in GFP<sup>+</sup> vs GFP<sup>-</sup> cells are shown. Each symbol represents one mouse. **b**, Relative expression of Vβ14 and Vβ6 TCRs by SILP TH17 versus non-TH17 CD4<sup>+</sup> T cells from *Il23r*<sup>GFP/+</sup> mice. Left, representative FACS plots; Right, analysis of multiple animals. **c**, Specific enrichment of Vβ14 TCRs in

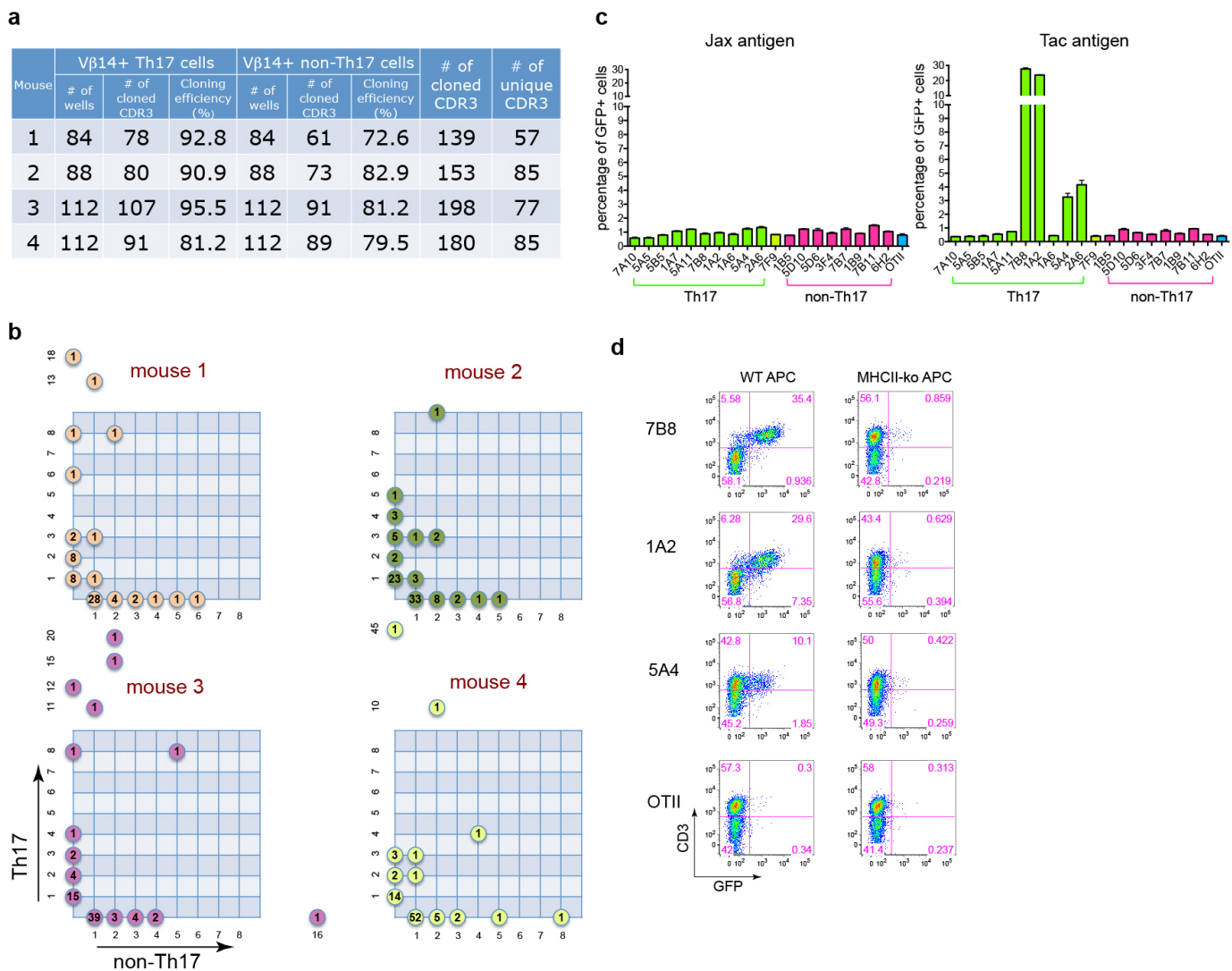
CD4<sup>+</sup> T cells expressing RORγt and IL-17A, but not FOXP3 or IFNγ. Left, representative FACS plots. Right, analysis of multiple animals. Each symbol represents one mouse. **d**, Correlation of Vβ14 enrichment in TH17 cells with the presence of specific commensal microbiota. B6 Jackson mice were housed alone or cohoused with B6 Taconic mice for two weeks. Left, representative FACS analyses. Right, analysis of multiple animals.





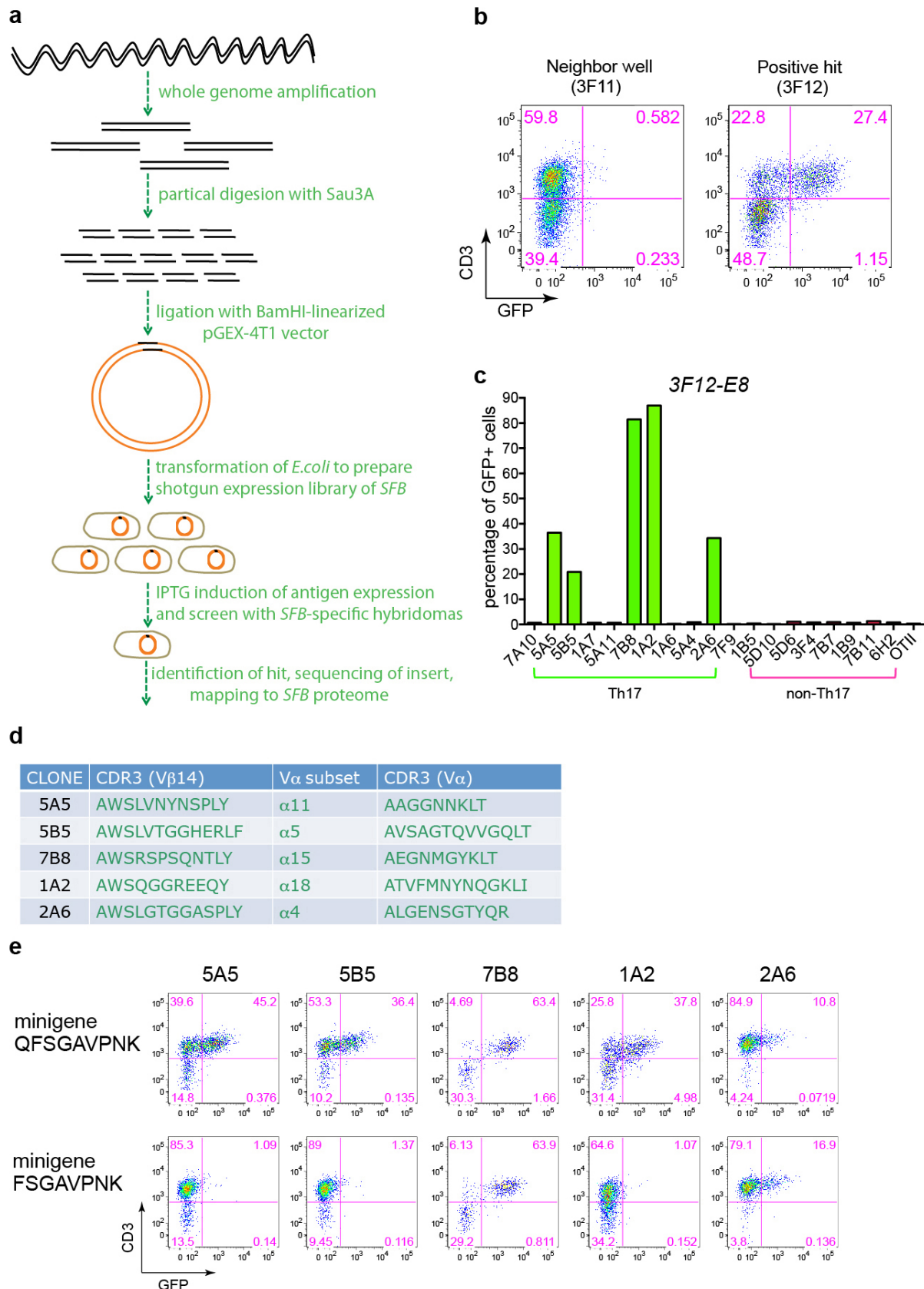
**Extended Data Figure 3 | T<sub>H</sub>17 TCR repertoire analysis by pyrosequencing.**  
**a**, Numbers of unique Vβ14 CDR3 sequences of individual SILP T<sub>H</sub>17 and non-T<sub>H</sub>17 samples. The sequences were normalized for numbers of cells and total reads. **b**, Preferential expansion of Vβ14<sup>+</sup> clones in the T<sub>H</sub>17

compartment in the SILP. The proportions of the 10 most abundant Vβ14 CDR3 sequences from T<sub>H</sub>17 and non-T<sub>H</sub>17 cells from 8 mice are shown.  
**c**, T<sub>H</sub>17-non T<sub>H</sub>17 bias of unique Vβ14 CDR3 sequences in the SILP of multiple mice.



**Extended Data Figure 4 | Single-cell TCR cloning and TCR hybridoma screen.** **a**, Efficiency of single-cell Vβ14 cloning from SILP TH17 and non-TH17 cells of multiple mice. **b**, Distributions of unique Vβ14 sequences in TH17 and non-TH17 cells within the SILP. Each plot represents one mouse shown in **a**. *y* and *x* axes represent numbers of TH17 cells and non-TH17 cells for

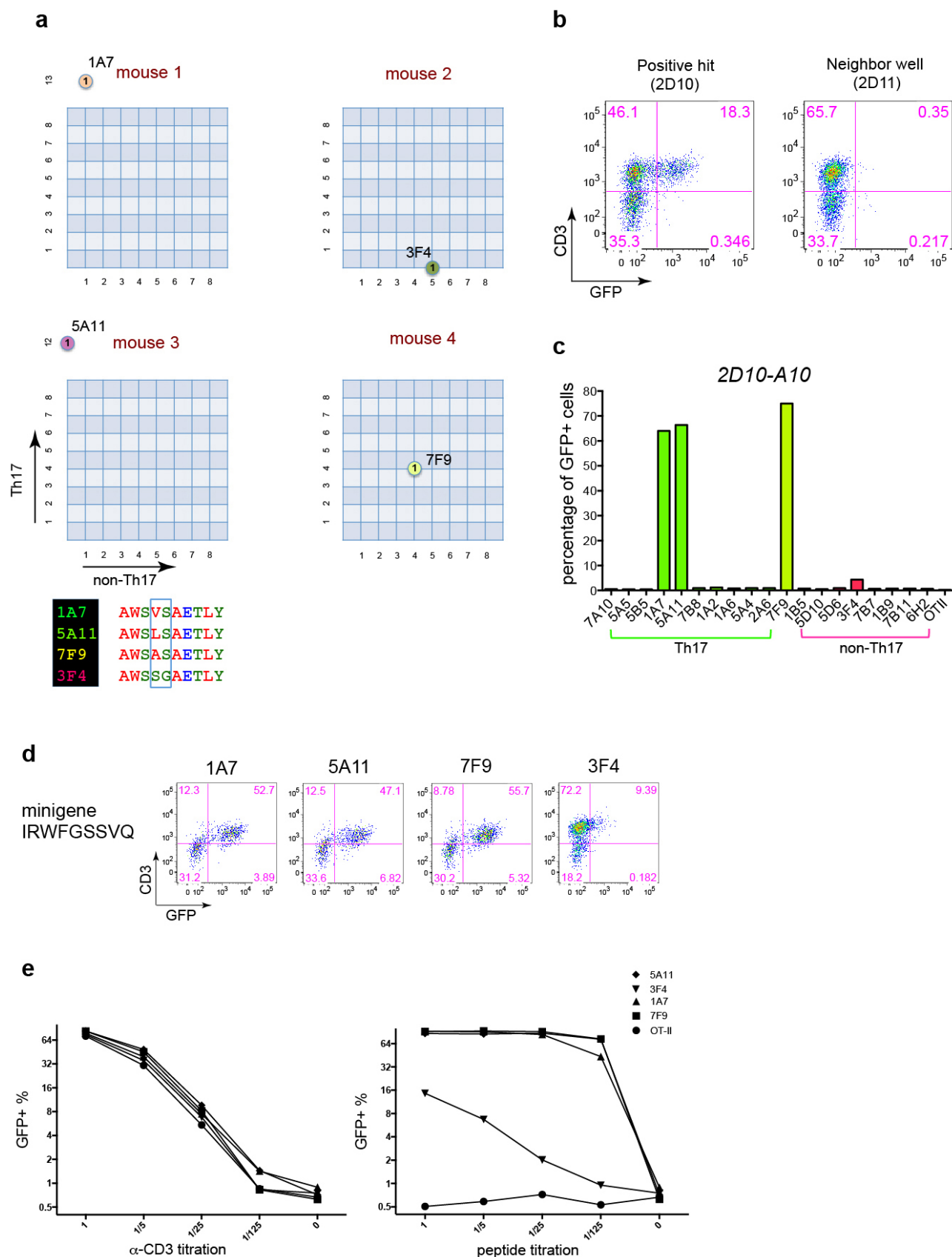
each unique Vβ14 sequence. Numbers of unique sequences are shown in coloured circles. **c**, Responses of TH17 and non-TH17 TCR hybridomas to small intestinal luminal contents from B6 Taconic and B6 Jackson mice. **d**, Stimulation of TH17 TCR hybridomas by SFB-monoassociated antigens in the presence of APCs sufficient (WT) or deficient (KO) for MHC class II.



**Extended Data Figure 5 | Identification of SFBNYU\_003340 epitopes recognized by a subset of the T<sub>H</sub>17 TCR hybridomas.** **a**, Schematic representation of the antigen screen using a whole-genome shotgun SFB library. **b**, Stimulation of the 7B8 hybridoma by bacterial pool 3F12. **c**, Reactivity of 7B8 and four other TCR hybridomas with bacterial clone

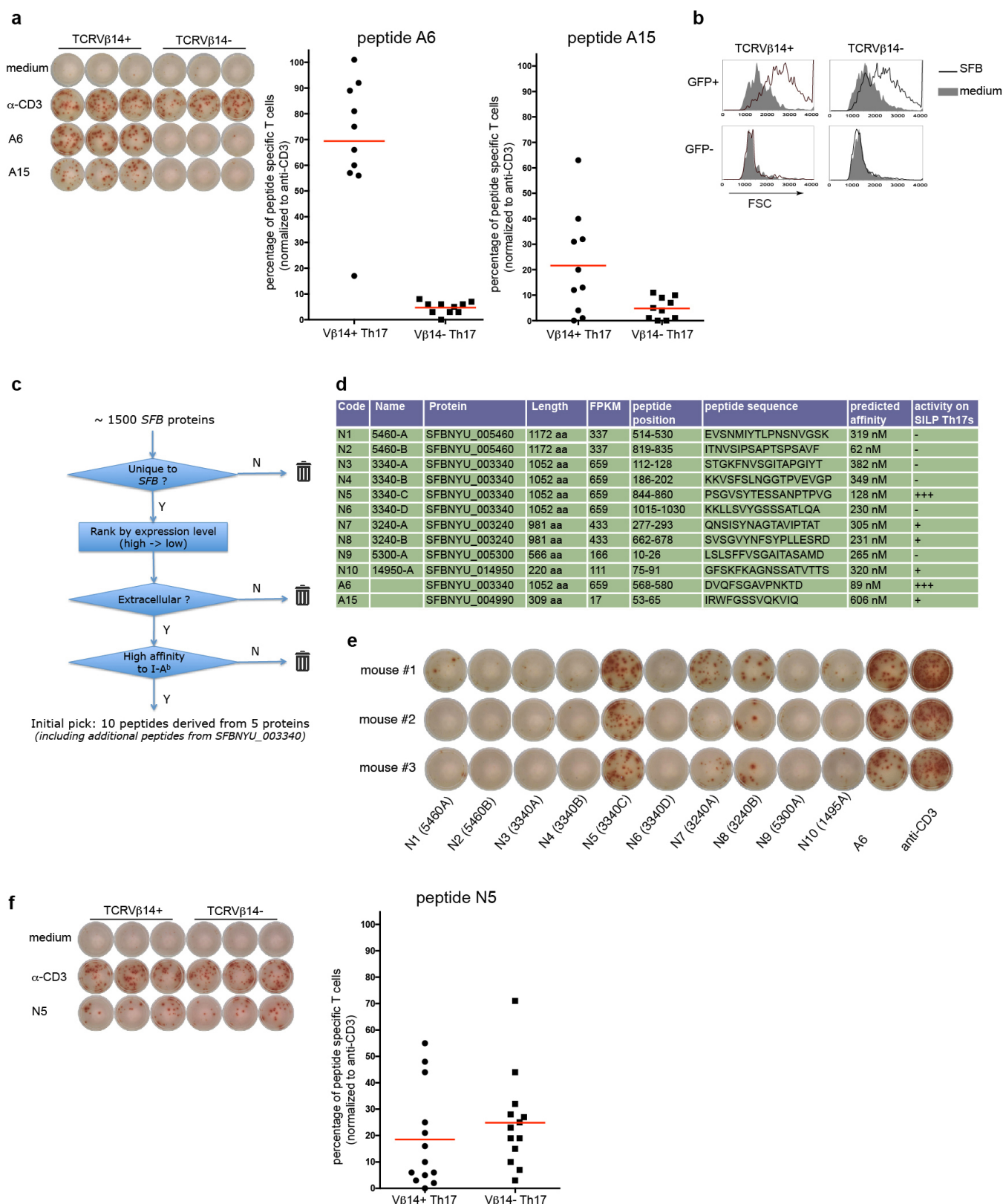
3F12-E8. **d**, Diversity of the CDR3 sequences of TCRs specific for 3F12-E8. Note that they belong to different V $\alpha$  subsets and have distinct V $\beta$ 14 CDR3 sequences. **e**, Responses of the 3F12-E8-specific TCR hybridomas to core epitopes encoded by minigenes expressed in *E. coli*.





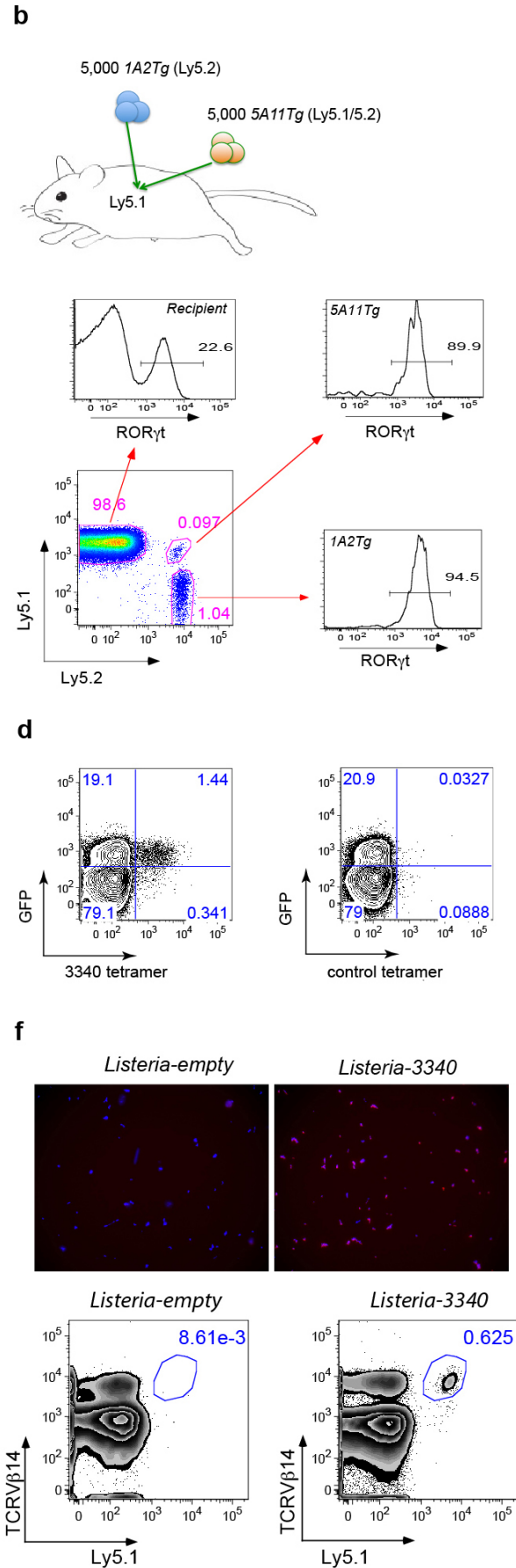
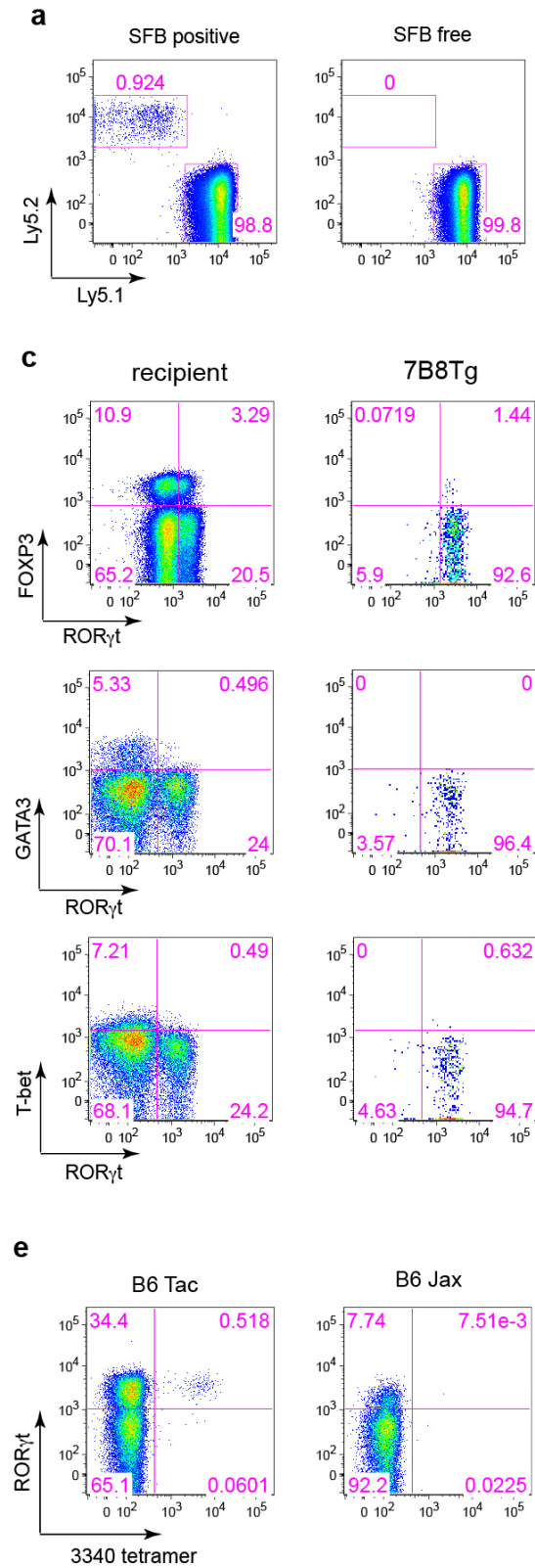
**Extended Data Figure 6 | Identification of *SFBNYU\_004990* epitopes recognized by related TCRs.** **a**, Top, the distribution in  $T_H17$  and non- $T_H17$  cells of four TCRs that share an identical TCR $\alpha$  chain. Bottom, amino acid alignment of the V $\beta$ 14 CDR3 sequences. The green box highlights the sequence differences. **b**, Stimulation of the 5A11 hybridoma by bacterial pool 2D10 in

the SFB antigen screen. **c**, Responses of 4 TCR hybridomas, including a non- $T_H17$  hybridoma, to bacterial clone 2D10-A10. **d**, Responses of the 2D10-A10-specific TCR hybridomas to core epitopes encoded by minigenes expressed in *E. coli*. **e**, TCR hybridoma responses to titrated synthetic peptide (IRWFGSSVQKV) in the presence of APCs.



**Extended Data Figure 7 | SFB epitopes recognized by diverse TH17 cell TCRs.** **a**, The epitopes recognized by the Vβ14<sup>+</sup> TCR hybridomas stimulate only Vβ14<sup>+</sup> TH17 cells from the SILP. TH17 cells sorted from *Il23r*<sup>GFP/+</sup> mice were stimulated with indicated peptides (listed in **d**) in the presence of APCs. Left, representative IL-17A ELISPOT assay with triplicates. Right, normalized peptide-specific TH17 responses. Each dot represents one mouse. **b**, Polyclonal responses of Vβ14<sup>+</sup> and Vβ14<sup>-</sup> SILP TH17 cells to SFB antigens. Representative FACS plots from five experiments are shown. **c**, Bioinformatics

filtering approach to select candidate SFB epitopes. **d**, Summary of newly selected and the known A6 and A15 SFB peptides. **e**, IL-17A ELISPOT screen for indicated peptides using SILP TH17 cells sorted from SFB-colonized *Il23r*<sup>GFP/+</sup> mice. The A6 peptide from *SFBNYU\_003340* and anti-CD3 served as positive controls. **f**, Vβ14 usage in TH17 cells specific for peptide N5. Left, representative IL-17A ELISPOT assay with triplicates for peptide N5, using Vβ14<sup>+</sup> and Vβ14<sup>-</sup> SILP TH17 cells sorted from *Il23r*<sup>GFP/+</sup> mice. Right, normalized N5-specific TH17 responses. Each dot represents one mouse.





**Extended Data Figure 8 | SFB-specific T cells become T<sub>H</sub>17 cells in**

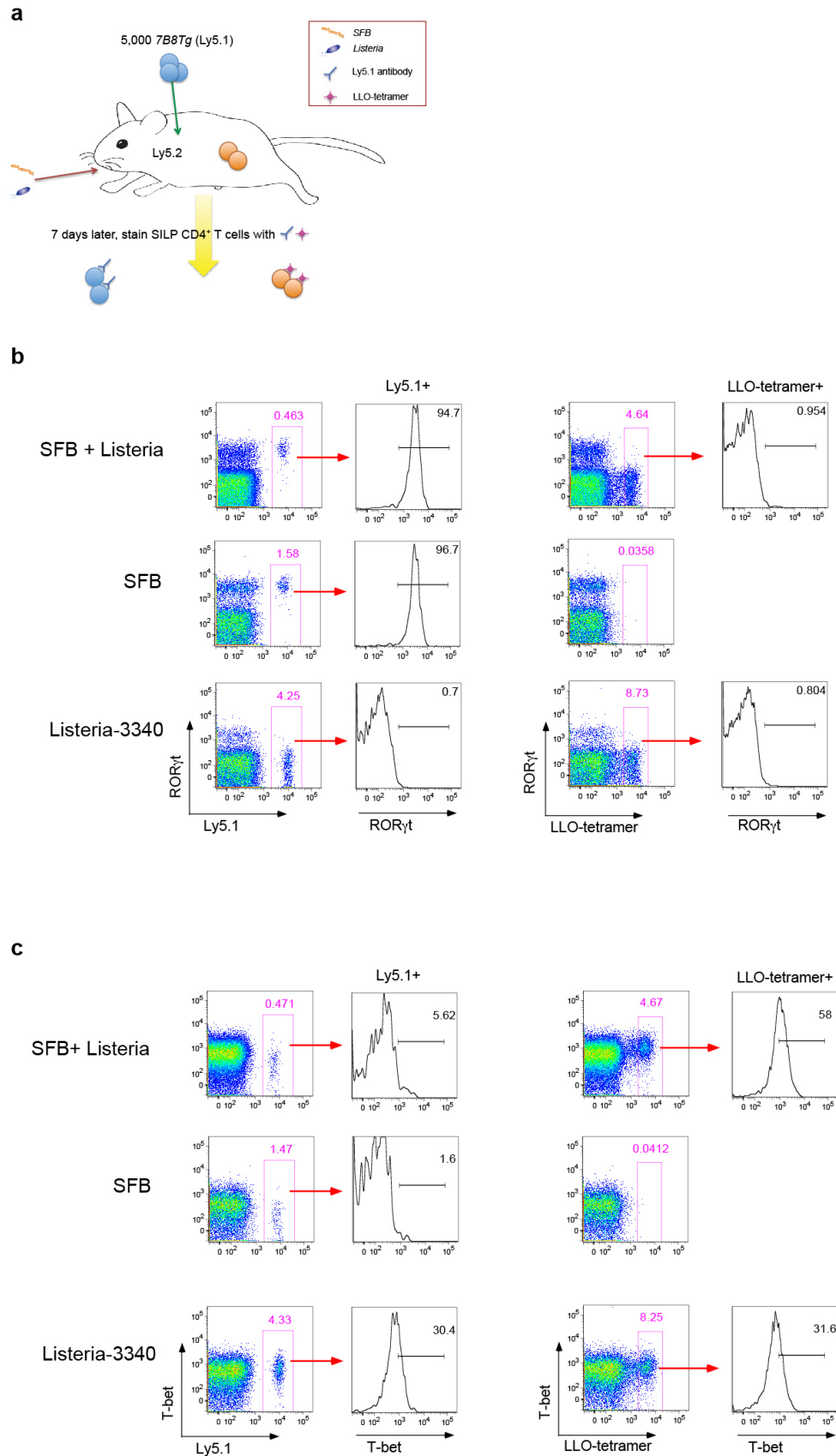
**SFB-colonized mice.** **a**, SFB-dependent 7B8Tg T cell accumulation in the SILP.  $2 \times 10^4$  naive 7B8Tg T cells were transferred into congenic Ly5.1 recipient mice that were SFB-colonized or SFB-free. CD4<sup>+</sup> T cells in the SILP were examined for donor and recipient isotype markers after 13 days.

**b**, Top, strategy for co-transfer of congenic 1A2Tg and 5A11Tg T cells into SFB-colonized recipient mice. Bottom, FACS analysis of ROR $\gamma$ t expression in host- and donor-derived CD4<sup>+</sup> T cells in the SILP at 7 days after transfer.

**c**, FACS analysis of transcription factors in host- and donor-derived SILP CD4<sup>+</sup> T cells after transfer of naive 7B8Tg T cells as in **a**. **d**, FACS analysis of SILP T cells from *Il23r<sup>GFP/+</sup>* mice, stained with I-A<sup>b</sup>/3340-A6 tetramer and

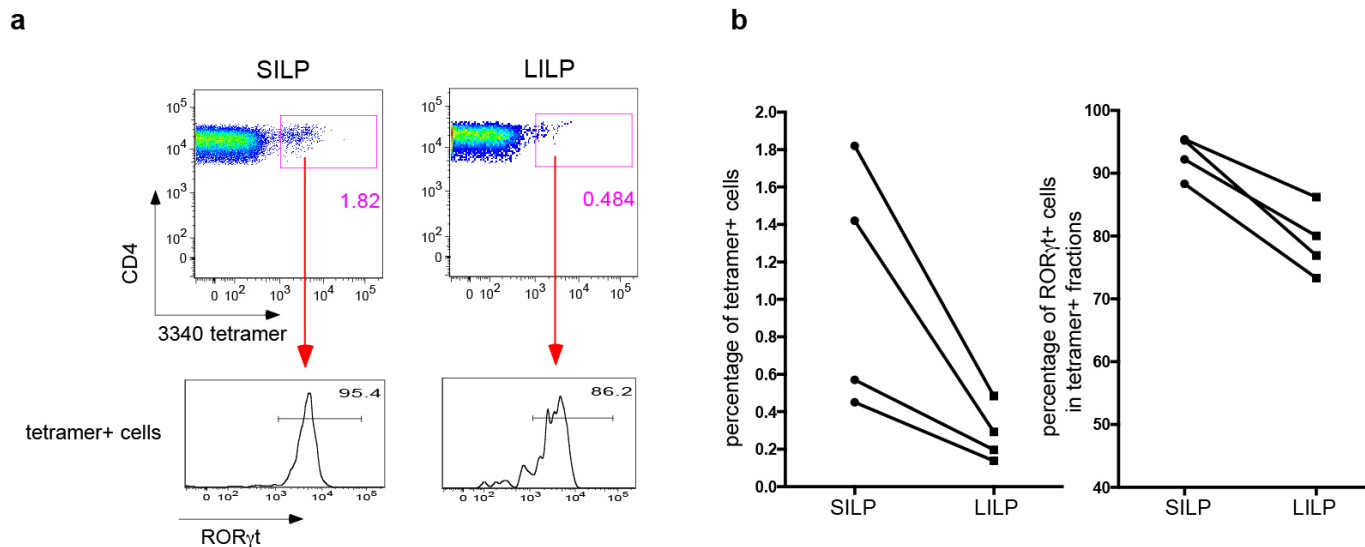
control tetramer (2W). **e**, FACS analysis of SILP T cells of B6 mice from colonies with different microbiota, stained with I-A<sup>b</sup>/3340-A6 tetramer and intracellular ROR $\gamma$ t antibody. **f**, Expansion of 7B8Tg T cells in mice colonized with *Listeria monocytogenes* expressing *SFB<sub>NYU</sub>\_003340*.

Top, immunofluorescence microscopic visualization of the expression of SFB protein by *L. monocytogenes*. *Listeria-3340* and *Listeria-empty* were stained with anti-3340 rabbit polyclonal antibody. Red, anti-3340 antibody staining. Blue, DAPI staining. Bottom, naive Ly5.1<sup>+</sup> 7B8Tg cells were transferred into congenic mice infected with *Listeria-3340* or *Listeria-empty*. Seven days after transfer, donor-derived CD4<sup>+</sup> T cells in the SILP were analysed by FACS.



**Extended Data Figure 9 | Transcription factor expression in SFB-specific and *Listeria*-specific T cells in co-infected mice.** Representative of data plotted in Fig. 4b. **a**, Experimental design for tracking both SFB- and *Listeria*-specific CD4<sup>+</sup> T cells following intestinal colonization with both bacteria. Ly5.2 B6 mice were colonized with *Listeria monocytogenes*, SFB, or

both bacteria, and 7B8Tg T cells from Ly5.1 mice were injected intravenously. Expression of T<sub>H</sub>1 and T<sub>H</sub>17 transcription factors in the SFB-specific 7B8Tg cells and LLO tetramer-specific recipient T cells was evaluated. **b**, Intracellular stain for RORγt. **c**, Intracellular stain for T-bet.



**Extended Data Figure 10 | SFB-specific T<sub>H</sub>17 cells are present in both SILP and large intestine lamina propria (LILP) of SFB-colonized mice.** T cells were stained with I-A<sup>b</sup>/3340-A6 tetramer and antibody to intracellular ROR $\gamma$ t. **a**, Representative FACS plots (gated on CD4<sup>+</sup> T cells). **b**, Analysis of multiple

animals. Left, per cent of tetramer-positive cells among total CD4<sup>+</sup> T cells in each region of the intestine. Right, per cent of ROR $\gamma$ t<sup>+</sup> cells among the tetramer-positive cells. Each symbol represents cells from a separate animal.



# Nociceptive sensory neurons drive interleukin-23-mediated psoriasiform skin inflammation

Lorena Riol-Blanco<sup>1\*</sup>, Jose Ordoñas-Montanes<sup>1\*</sup>, Mario Perro<sup>1</sup>, Elena Naval<sup>1</sup>, Aude Thiriot<sup>1</sup>, David Alvarez<sup>1</sup>, Silke Paust<sup>1†</sup>, John N. Wood<sup>2</sup> & Ulrich H. von Andrian<sup>1</sup>

**The skin has a dual function as a barrier and a sensory interface between the body and the environment. To protect against invading pathogens, the skin harbours specialized immune cells, including dermal dendritic cells (DDCs) and interleukin (IL)-17-producing  $\gamma\delta$  T ( $\gamma\delta$ T17) cells, the aberrant activation of which by IL-23 can provoke psoriasis-like inflammation<sup>1–4</sup>. The skin is also innervated by a meshwork of peripheral nerves consisting of relatively sparse autonomic and abundant sensory fibres. Interactions between the autonomic nervous system and immune cells in lymphoid organs are known to contribute to systemic immunity, but how peripheral nerves regulate cutaneous immune responses remains unclear<sup>5,6</sup>. We exposed the skin of mice to imiquimod, which induces IL-23-dependent psoriasis-like inflammation<sup>7,8</sup>. Here we show that a subset of sensory neurons expressing the ion channels TRPV1 and Na<sub>v</sub>1.8 is essential to drive this inflammatory response. Imaging of intact skin revealed that a large fraction of DDCs, the principal source of IL-23, is in close contact with these nociceptors. Upon selective pharmacological or genetic ablation of nociceptors<sup>9–11</sup>, DDCs failed to produce IL-23 in imiquimod-exposed skin. Consequently, the local production of IL-23-dependent inflammatory cytokines by dermal  $\gamma\delta$ T17 cells and the subsequent recruitment of inflammatory cells to the skin were markedly reduced. Intradermal injection of IL-23 bypassed the requirement for nociceptor communication with DDCs and restored the inflammatory response<sup>12</sup>. These findings indicate that TRPV1<sup>+</sup>Na<sub>v</sub>1.8<sup>+</sup> nociceptors, by interacting with DDCs, regulate the IL-23/IL-17 pathway and control cutaneous immune responses.**

Repeated topical application of imiquimod (IMQ) to murine skin provokes inflammatory lesions that resemble human psoriasis<sup>7,8</sup>. This response is mediated by IL-23, which stimulates skin-resident  $\gamma\delta$  T cells to secrete IL-17 and IL-22, cytokines that induce inflammatory leukocyte recruitment and acanthosis<sup>13</sup>. Indeed, antibodies targeting the shared p40 subunit of IL-12 and IL-23 inhibit both IMQ-induced murine dermatitis and human psoriasis<sup>8,13</sup>. Frequent symptoms in human psoriasis, aside from the prominent skin lesions, include the sensations of itch, pain and discomfort in affected areas<sup>14</sup>. Clinical reports suggest that intralesionally administered anaesthetics or surgical denervation of psoriatic lesions not only abrogate local sensation, but also ameliorate local inflammation<sup>15</sup>. Similarly, in mutant mice with disseminated psoriasiform dermatitis, peripheral nerve dissection attenuated skin inflammation<sup>16</sup>; however, cutaneous nerves are composed of sympathetic and several types of sensory fibres, and the role of individual types of nerve fibres remains unclear<sup>6</sup>. Using the IMQ model, we investigated whether and how specific subsets of peripheral nerves contribute to the formation of psoriasiform skin lesions.

Skin sensations perceived as inflammatory pain, noxious heat and some forms of itch are transmitted by sensory fibres that express the cation channel TRPV1. Most TRPV1<sup>+</sup> fibres co-express the sodium channel Na<sub>v</sub>1.8 (refs 9–11). Na<sub>v</sub>1.8<sup>+</sup> nociceptors can be identified in the dermis of Na<sub>v</sub>1.8-TdTomato (TdT) mice by their red fluorescence<sup>9</sup> (Fig. 1a).

Confocal microscopy of skin samples from Na<sub>v</sub>1.8-TdT mice co-stained for tyrosine hydroxylase (TH), which identifies sympathetic fibres, and for  $\beta$ 3-tubulin, a pan-neuronal marker, revealed that Na<sub>v</sub>1.8<sup>+</sup>TH<sup>+</sup> nociceptors represent the vast majority of cutaneous nerve fibres, whereas Na<sub>v</sub>1.8<sup>+</sup>TH<sup>+</sup> sympathetic fibres are rare.

To investigate the roles of sympathetic fibres and nociceptors in the IMQ model, mice were treated systemically with either 6-hydroxydopamine (6OHDA) or resiniferatoxin (RTX) to ablate TH<sup>+</sup> sympathetic neurons or TRPV1<sup>+</sup> nociceptors, respectively<sup>11,17</sup> (Extended Data Figs 1, 2). Subsequently, IMQ was applied topically to one ear and the ensuing inflammatory response was assessed based on the change in ear thickness, size of the myeloid infiltrate (Extended Data Fig. 3a) and tissue contents of inflammatory cytokines.

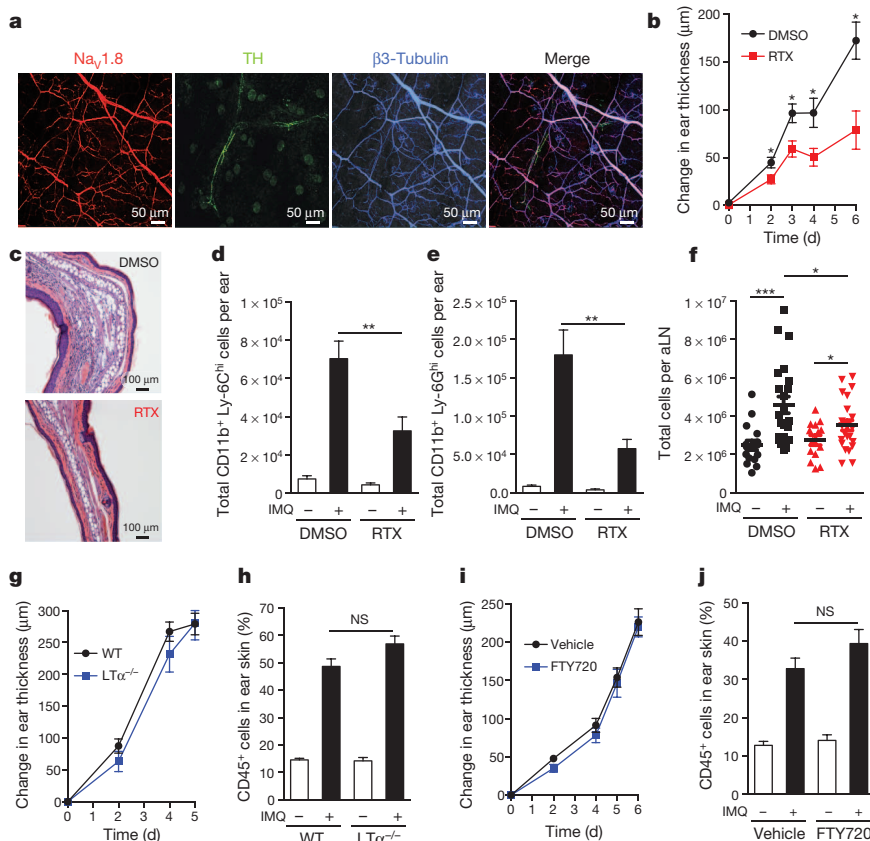
After sympathetic denervation, IMQ-induced ear swelling was reduced compared to controls (Extended Data Fig. 1c); however, the inflammatory infiltrate was increased, whereas IL-17A, IL-17F, IL-22 and IL-12/23p40 production remained unchanged (Extended Data Fig. 1d–i). Thus, sympathetic innervation exerts little or no direct local control over the inflammatory skin response. The observed changes were probably due to cardiovascular effects and/or global immune dysregulation after systemic sympathectomy<sup>5</sup> (Supplementary Information). By contrast, in RTX-treated mice both ear swelling and inflammatory infiltrates in IMQ-exposed ears were profoundly reduced (Fig. 1b–e and Extended Data Fig. 4a–d). RTX treatment did not alter the systemic supply of inflammatory cells<sup>18</sup> (Extended Data Fig. 4e, f). Moreover, intravital microscopy of ear skin revealed similar leukocyte rolling in RTX-treated and control mice (Extended Data Fig. 5), indicating that the absence of nociceptors did not affect the baseline adhesiveness of dermal microvessels<sup>19</sup>. It is more likely that ablation of TRPV1<sup>+</sup> sensory nerves reduced IMQ-induced inflammation through local, extravascular mechanisms. However, the attenuated inflammatory response was not limited to the skin, as the IMQ-induced enhancement in cellularity of the draining auricular lymph node was also blunted by RTX (Fig. 1f).

Lymph nodes have a critical function in dermal antigen presentation to naive T cells and in the generation of migratory effector cells. They also possess peripheral innervation<sup>20</sup>, which could have been altered by RTX; however, the role of skin-draining lymph nodes during psoriatic inflammation is unclear. To address this issue, we tested the effect of IMQ in lymphotoxin- $\alpha$ -deficient (LT $\alpha$ <sup>−/−</sup>) mice, which are devoid of lymph nodes. Compared with wild-type mice, there was no statistical difference in ear thickness, frequency or composition of the inflammatory infiltrate in ears of LT $\alpha$ <sup>−/−</sup> mice (Fig. 1g, h and data not shown). Additionally, we treated wild-type mice with FTY720, which blocks T-cell egress from lymph nodes, preventing trafficking of effector cells to peripheral tissues<sup>21</sup>. Again, IMQ elicited full-fledged inflammation in exposed ears (Fig. 1i, j and data not shown), indicating that T-cell priming in skin draining lymph nodes is dispensable for the acute induction of psoriasiform inflammation. Together, these findings imply that TRPV1<sup>+</sup> nociceptors promote local immune responses directly in the skin.

<sup>1</sup>Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>2</sup>Institute for Biomedical Research, University College London, London WC1E 6BT, UK.

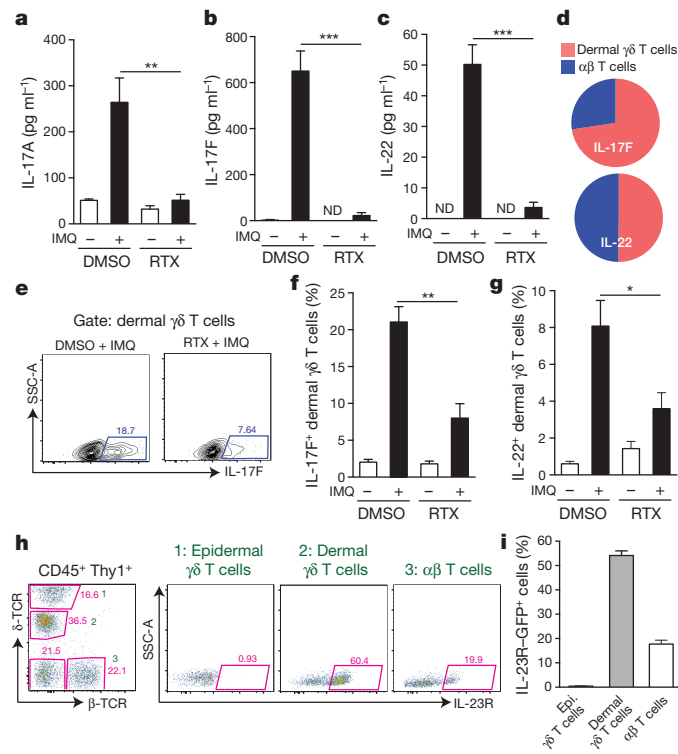
<sup>†</sup>Present address: Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA.

\*These authors contributed equally to this work.



**Figure 1 | TRPV1<sup>+</sup> nociceptor ablation attenuates skin inflammation and draining lymph node hypertrophy in the IMQ model.** **a**, Representative whole-mount confocal micrograph of normal ear skin from Na<sub>v</sub>1.8-TdT reporter mice (Na<sub>v</sub>1.8<sup>+</sup> nociceptors, red) stained for TH (sympathetic nerves, green) and β3-tubulin (peripheral nerves, blue). **b–f**, The ear skin of vehicle (dimethylsulphoxide (DMSO))-treated controls or TRPV1<sup>+</sup> nociceptor ablated (RTX) mice was treated with topical IMQ cream daily. **b**, Ear thickness was measured relative to the contralateral ear at indicated time points ( $n = 10–15$  mice per time point;  $*P < 0.02$ ). **c**, Representative histological sections of IMQ-treated ears at day 6 stained by haematoxylin and eosin. **d, e**, Total inflammatory monocytes ( $n = 10$ ) (**d**) and total neutrophils (**e**) in skin at day 3 ( $n = 10$ ;  $**P < 0.005$ ). **f**, Total cell number in auricular lymph nodes (aLNs) at day 3 ( $n = 20$ ;  $*P = 0.01$ ;  $***P < 0.001$ ). **g–j**, IMQ was applied daily to wild-type (WT;  $n = 10$ ) and LTα<sup>-/-</sup> mice ( $n = 6$ ) (**g, h**) or vehicle-treated ( $n = 10$ ) and FTY720-treated mice ( $n = 10$ ) (**i, j**), and ear swelling was measured at the indicated time points (**g, i**). **h, j**, The percentage of CD45<sup>+</sup> leukocytes was determined in ear skin digests on day 6. NS, not significant. Error bars show mean  $\pm$  standard error of the mean (s.e.m.).

Neutrophil recruitment to the skin and keratinocyte hyperproliferation, hallmarks of IMQ-induced inflammation, are driven by IL-17 and IL-22, respectively<sup>2,7,12</sup>. We asked whether nociceptors regulate the production of these cytokines. Indeed, after IMQ treatment, protein levels of IL-17A, IL-17F and IL-22 had markedly increased in the ears of control mice, whereas in RTX-treated mice IL-17A was very low and IL-17F and IL-22 remained below the detection limit (Fig. 2a–c). Thus,



TRPV1<sup>+</sup> nociceptors control the generation of several key effector cytokines in psoriasiform dermatitis.

As our findings in FTY720-treated mice suggested that recruitment of lymph-node-derived effector T cells is not needed for IMQ-induced skin inflammation, we asked whether nociceptors regulate cytokine production by skin-resident lymphocytes, particularly γδT17 cells, which produce IL-17F and IL-22 (refs 3, 22, 23). In the ears of IMQ-treated control mice, most IL17F<sup>+</sup> lymphocytes were dermal γδ T cells, whereas few conventional αβ T cells expressed IL-17F (Fig. 2d and Extended Data Figs 3b, 6a). In RTX-treated mice, both IL-17F<sup>+</sup> and IL-22<sup>+</sup> dermal γδ T cells were significantly reduced, indicating that TRPV1<sup>+</sup> neurons drive the local production of IL-17F and IL-22 primarily by γδT17 cells (Fig. 2e–g and Extended Data Fig. 6b, c). These T cells seed the skin in early life and are poised for rapid IL-17 production upon stimulation by IL-23 (ref. 24). Indeed, in the ears of IL-23R<sup>GFP/+</sup> mice<sup>25</sup>, ~60% of dermal γδT17 and ~20% of αβ T cells expressed IL-23R (Fig. 2h, i), whereas T cells in auricular lymph nodes and epidermal γδ T cells expressed little or no IL-23R (Extended Data Fig. 6d, e).

In light of the preferential expression of IL-23R on dermal γδT17 cells, the known role of IL-23 as a driver of IL-17 and IL-22 generation

**Figure 2 | TRPV1<sup>+</sup> nociceptors control IL-17F and IL-22 production by IL-23R<sup>+</sup> dermal γδ T cells.** **a–c**, After 3 days of IMQ challenge, ears from vehicle (DMSO)-treated mice ( $n = 5$ ) or RTX-treated mice ( $n = 5$ ) were harvested to perform enzyme-linked immunosorbent assays (ELISA) for IL-17A (**a**), IL-17F (**b**) and IL-22 (**c**).  $**P < 0.01$ ;  $***P < 0.001$ . ND, not detected. **d–g**, Flow cytometry on digested ear skin was performed on day 6 of IMQ challenge. **d**, Relative frequency of IL-17F<sup>+</sup> or IL-22<sup>+</sup> dermal γδ T cells and αβ T cells in IMQ-treated control mice ( $n = 15$ ). **e**, Representative FACS plots of IL-17F staining in dermal γδ T cells from vehicle (DMSO)- and RTX-treated mice. **f, g**, Quantification of frequency of IL-17F<sup>+</sup> (**f**) and IL-22<sup>+</sup> (**g**) among dermal γδ T cells in DMSO- and RTX-treated mice ( $n = 5$  per group;  $*P < 0.05$ ;  $**P = 0.01$ ). **h, i**, Representative FACS plots of normal ear skin from an IL-23R<sup>GFP/+</sup> mouse (**h**) and quantification of frequency of IL-23R-GFP<sup>+</sup> cells among skin-resident Thy1<sup>+</sup> T-cell subsets ( $n = 8$ ) (**i**). Epi., epidermal. Error bars show mean  $\pm$  s.e.m.

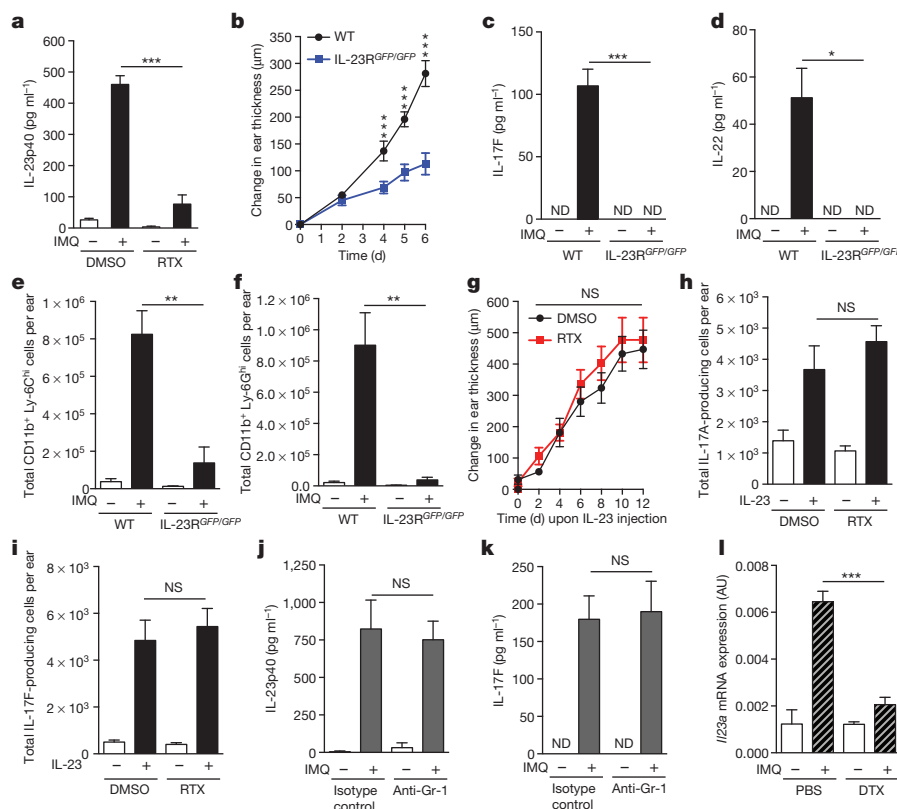
in psoriasiform inflammation<sup>13</sup> and our finding that RTX pre-treatment profoundly diminished IMQ-induced cytokine production by  $\gamma\delta T17$  cells, we proposed that TRPV1<sup>+</sup> nociceptors might control dermal IL-23 production. Topical IMQ treatment of control mice markedly increased p40 protein levels, as well as messenger RNA levels of *Il12b* (IL-12/23p40) and *Il23a* (IL-23p19), but not *Il12a* (IL-12p35). These effects were nearly abolished after RTX treatment (Fig. 3a and Extended Data Fig. 7a–c), suggesting that TRPV1<sup>+</sup> nociceptors are essential for cutaneous IL-23 production. It seems unlikely that IL-12 had a major impact in the IMQ model, because IMQ-induced inflammatory parameters were considerably reduced in IL-23R<sup>GFP/GFP</sup> mice, which respond to IL-12 but not IL-23 (Fig. 3b–f). However, IL-12-dependent skin inflammation induced using a chemical irritant, 2,4-dinitrofluorobenzene (DNFB)<sup>26</sup>, was profoundly reduced in RTX-treated mice, suggesting that RTX-sensitive fibres also have a role in IL-12-driven dermatitis (Extended Data Fig. 7d, e).

Of note, although IL-17F and IL-22 production as well as myeloid infiltrates were virtually abolished in IL-23R<sup>GFP/GFP</sup> mice, ear swelling was only partially reduced (Fig. 3b–f), suggesting that topical IMQ promotes modest tissue oedema through an IL-23-independent pathway. This activity was independent of nociceptors, as RTX treatment of IL-23R<sup>GFP/GFP</sup> mice had no effect on the IMQ-induced swelling (Extended Data Fig. 7f).

Although the above findings clearly demonstrate that nociceptors are indispensable for IMQ-induced dermal IL-23 production, it remained

possible that nociceptive fibres exert additional pro-inflammatory functions, for example, by directly regulating  $\gamma\delta T17$  cells. To address this, we performed intradermal IL-23 injections, which trigger psoriasiform skin inflammation in murine skin<sup>12</sup>. Regardless of whether nociceptors were ablated or left intact, IL-23 administration resulted in profound ear swelling and fully rescued IL-17A and IL-17F production by Thy1<sup>+</sup> cells (Fig. 3h, i) and  $\gamma\delta T17$  cells (Extended Data Fig. 7g, h). Together, these results suggest that the pro-inflammatory function of TRPV1<sup>+</sup> neurons is rooted exclusively in the promotion of IL-23 production, at least in this experimental setting.

Next, we sought to identify the IL-23-producing cell type(s) regulated by TRPV1<sup>+</sup> nociceptors. In intestinal barrier tissues, Ly-6C<sup>high</sup> inflammatory monocytes were identified as a major source of IL-23 in a colitis model<sup>27</sup>. However, depletion of neutrophils and inflammatory monocytes with anti-Gr-1 did not affect IMQ-induced dermal IL-23p40 or IL-17F levels (Fig. 3j, k and Extended Data Fig. 8a), suggesting that skin-resident dendritic cells and/or macrophages rather than migratory myeloid cells were the critical source of IL-23. Thus, we injected CD11c-diphtheria toxin receptor (DTR) mice with diphtheria toxin (DTX), which depleted DDCs and Langerhans cells, but not macrophages (Extended Data Fig. 8b, c). After DTX treatment, IMQ-induced expression of *Il23a* mRNA was markedly reduced in treated ears (Fig. 3l). Although these results do not discriminate between the relative contributions of Langerhans cells and DDCs to IL-23 production, IL-34-deficient mice—which



**Figure 3 | DDC-derived IL-23 is critical to drive psoriasiform skin inflammation and acts downstream of RTX-sensitive nociceptors.** **a**, After 3 days of IMQ challenge of vehicle (DMSO)- or RTX-treated mice, ears were harvested and total protein was prepared to quantify IL-23p40 by ELISA ( $n = 5$  per experiment; \*\*\* $P < 0.001$ ). **b**, Ears of wild-type (WT) or IL-23R<sup>GFP/GFP</sup> mice ( $n = 5$  per group) were treated daily with IMQ and ear thickness was measured relative to the contralateral ear at the indicated time points (\*\*\* $P < 0.001$ ). **c–f**, After 3 days of IMQ challenge in wild-type ( $n = 5$ ) or IL-23R<sup>GFP/GFP</sup> mice ( $n = 4$ ) total protein was prepared from ear skin and IL-17F (**c**) and IL-22 (**d**) were quantified by ELISA (\* $P < 0.05$ ; \*\*\* $P < 0.001$ ). Cell suspensions from exposed ears stained for inflammatory monocytes (**e**) and neutrophils (**f**) to assess total numbers ( $n = 5$  per experiment; \* $P < 0.05$ ; \*\* $P < 0.01$ ). **g–i**, IL-23 was injected intradermally every other day into the ears

of DMSO- or RTX-treated mice ( $n = 8$  per group). **g**, Ear thickness was measured as indicated. **h, i**, After 3 days, the number of IL-17A-producing (**h**) and IL-17F-producing (**i**) Thy1<sup>+</sup> cells per ear was quantified by FACS ( $n = 5$ ). **j, k**, Mice were treated with anti-Gr-1 to deplete neutrophils and inflammatory monocytes or isotype-matched control monoclonal antibody and challenged with IMQ for 3 days. Ear skin protein lysates were analysed for IL-23p40 (**j**) and IL-17F (**k**) by ELISA ( $n = 5$ ). **l**, CD11c-DTR mice were treated with DTX or PBS 12 h before IMQ challenge. After 6 h, ears were harvested and processed for total RNA isolation and *Il23a* mRNA levels were analysed by quantitative polymerase chain reaction (qPCR) ( $n = 4$ ; \*\*\* $P = 0.001$ ). AU, arbitrary units. ND, not detected; NS, not significant. Error bars show mean  $\pm$  s.e.m.



selectively lack Langerhans cells—are not impaired in IMQ-induced skin inflammation<sup>8</sup>. We focused our analysis on DDCs, which can be further subdivided into CD103<sup>+</sup> and CD11b<sup>+</sup> subsets<sup>28</sup>. Both subsets, as well as macrophages, were sorted by fluorescence-activated cell sorting (FACS) from IMQ-treated and control ears to measure mRNA for *Il23a* and *Il12a*. Although CD103<sup>+</sup> DDCs showed the most notable upregulation of *Il23a* mRNA, taking into consideration that CD11b<sup>+</sup> DDCs are more abundant, we estimate that the latter subset produced ~75% of the *Il23a* mRNA, consistent with a recent report<sup>29</sup> (Fig. 4a and Extended Data Fig. 8c, d).

Having identified DDCs as the principal source of IMQ-induced IL-23, we sought to characterize the spatial relationship between DDCs and cutaneous nerves. Remarkably, confocal microscopy of skin whole mounts revealed that at steady state ~75% of DDCs were either in direct contact or in close proximity to sensory nerves (Fig. 4b–d and Extended Data Fig. 9a). Interactions were apparent along the entire length of nerves, suggesting that DDCs may receive signals from unmyelinated nociceptor axons and not just from nerve terminals. However, given the high density of peripheral nerves in the skin, it was difficult to judge whether the association with DDCs occurred merely by chance or reflected a biased distribution. To address this possibility, we compared DDC localization relative to two other dense anatomical structures: blood and lymph vessels. In resting tissues, contacts of DDCs with these microvascular networks were only about half as frequent as with peripheral nerves (Fig. 4d).

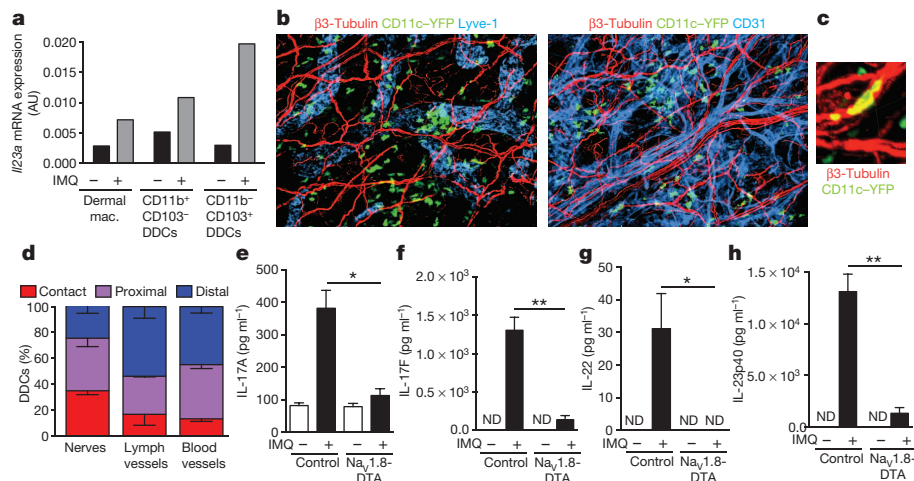
Of note, even though inflammation enhances DDC motility and egress into draining lymphatics, IMQ challenge did not alter the frequency of DDCs contacting nerves (Extended Data Fig. 9b), suggesting that DDCs engaged in dynamic interactions with nociceptors. We reconstituted Na<sub>v</sub>1.8-TdT mice with yellow fluorescent protein (YFP)-tagged CD11c bone marrow and performed time-lapse multiphoton-intravital microscopy (MP-IVM) in ear skin to generate three-dimensional time-lapse videos of interactions between YFP<sup>+</sup> DDCs and TdT<sup>+</sup> nociceptors. These interactions were highly diverse (Supplementary Video 1); some DDCs seemed to be anchored on nerves and sometimes extended protrusions to probe the surrounding tissue, whereas others appeared to use nerve fibres as a scaffold for directional migration (Supplementary Video 2).

Together, our results support the idea that DDCs can physically interact with a subset of nociceptors that regulate the production of IL-23.

However, it should be cautioned that the pharmacological target of RTX, TRPV1, is also expressed on some non-neuronal cells<sup>10</sup>. Consequently, we sought an alternative to confirm the role of nociceptors in an experimental system that does not rely on pharmacologically targeting TRPV1. To this end, we used Na<sub>v</sub>1.8-diphtheria toxin (DTA) mice in which Na<sub>v</sub>1.8<sup>+</sup> fibres, which respond to mechanical pressure, inflammatory pain and noxious cold, are selectively deleted<sup>9</sup>. Although nociceptor-associated transcripts in dorsal root ganglia were more profoundly reduced in this genetic model than after RTX treatment, *Trpv1* mRNA levels were only reduced by ~80%, consistent with the fact that noxious heat sensing fibres express TRPV1, but not Na<sub>v</sub>1.8 (Extended Data Figs 2c, 9c)<sup>9</sup>. After IMQ challenge, the ears of Na<sub>v</sub>1.8-DTA mice contained very low protein levels of IL-17A, IL-17F, IL-22 and IL-23p40 as compared to littermate controls (Fig. 4e–h). We conclude that Na<sub>v</sub>1.8<sup>+</sup> TRPV1<sup>+</sup> neurons are insufficient to induce IL-23 production; rather, Na<sub>v</sub>1.8<sup>+</sup> TRPV1<sup>+</sup> nociceptors are driving the response.

In light of these results, we propose a model of cutaneous neuroimmune interactions (Extended Data Fig. 10) whereby dermal Na<sub>v</sub>1.8<sup>+</sup> TRPV1<sup>+</sup> nociceptors are essential to induce IL-23 production by nearby DDCs. IL-23 then acts on IL-23R<sup>+</sup> γδT17 cells to induce IL-17F and IL-22 secretion, which precipitates the recruitment of circulating neutrophils and monocytes driving psoriasiform skin inflammation. The fact that both RTX and Na<sub>v</sub>1.8-DTA mice largely preserve a dense meshwork of dermal nerves (Extended Data Fig. 9d, e) implies that DDCs do not simply rely on nerves as a scaffold from which to produce IL-23. It is more likely that Na<sub>v</sub>1.8<sup>+</sup> TRPV1<sup>+</sup> nociceptors actively induce and control IL-23 production.

Although further studies will be needed to determine the precise molecular underpinnings of neuroimmune communication in the skin (Supplementary Information), the present findings indicating nociceptor-mediated control of DDCs and the IL-23/IL-17 axis open new avenues for the treatment of inflammatory diseases in the skin and perhaps elsewhere. Intriguingly, recent work has shown that Na<sub>v</sub>1.8<sup>+</sup> nerve fibres exert immunosuppressive activity during *Staphylococcus aureus* infection<sup>30</sup>, raising the possibility that some pathogens have evolved mechanisms to subvert the pro-inflammatory function of nociceptors. Together, this recent work and the present study suggest an emerging new paradigm whereby TRPV1<sup>+</sup> Na<sub>v</sub>1.8<sup>+</sup> nociceptive fibres integrate environmental



**Figure 4 | DDCs are closely associated with cutaneous nerves and depend on Na<sub>v</sub>1.8<sup>+</sup> nociceptors for IMQ-induced IL-23 production.** **a**, Mice were challenged with IMQ ( $n = 20$  pooled mice per condition) and, after 6 h, myeloid cell populations comprising dermal macrophages (mac.) and two subsets of DDCs were FACS sorted (Extended Data Fig. 8c) from cell suspensions to measure *Il23a* mRNA by qPCR. **b**, Representative confocal micrographs of ear skin whole mounts from CD11c-YFP mice stained for β3-tubulin (peripheral nerves, red) and Lyve-1 (lymphatics, blue) or CD31 (blood and lymphatic endothelial cells, blue). Original magnification, ×200. **c**, Close-up confocal micrograph of a CD11c-YFP cell in contact with

a nerve (see also Supplementary Videos 1, 2). **d**, Quantification of three-dimensional DDC proximity to peripheral nerves, lymphatics and blood vessels in normal ear skin. The frequency of DDCs ( $n = 330$ ) in contact, proximal (0–7 μm) and distal (>7 μm) to nerve fibres was determined as described in Methods and a Chi-squared test showed bias of DDCs to nerves relative to lymphatics and blood vessels (\*\* $P < 0.0001$ ). **e–h**, Ears of Na<sub>v</sub>1.8-DTA or control littermates were treated daily with IMQ. Total protein was prepared from ear skin after 3 days and IL-17A (**e**), IL-17F (**f**), IL-22 (**g**) and IL-23p40 (**h**) were quantified by ELISA ( $n = 4$  ears per condition; \* $P < 0.05$ ; \*\* $P < 0.01$ ). ND, not detected. Error bars show mean ± s.e.m.

signals to modulate local immune responses to a variety of infectious and inflammatory stimuli.

## METHODS SUMMARY

Published mouse strains used in this study are detailed in Methods. Na<sub>v</sub>1.8-Cre mice<sup>9</sup> were bred with Rosa26-DTA and Rosa26-TdT mice to generate Na<sub>v</sub>1.8-DTA mice for functional studies and Na<sub>v</sub>1.8-TdT mice for imaging, respectively. TRPV1<sup>+</sup> nociceptors were deleted using three escalating doses (30 µg kg<sup>-1</sup>, 70 µg kg<sup>-1</sup> and 100 µg kg<sup>-1</sup>) of RTX as described previously<sup>11</sup>. To induce psoriasiform ear inflammation, 8–12-week-old mice were treated topically with 5% IMQ cream or injected with 500 ng per ear recombinant (r)IL-23. Ear thickness was measured using an engineer's micrometer (Mitutoyo). Cytokines were quantified from skin protein extracts by ELISA (Biolegend, R&D). For flow cytometric analysis of tissue leukocyte markers and intracellular cytokines, single-cell suspensions from ear skin were prepared by enzymatic digestion<sup>4</sup>. Imaging of fixed skin tissue was performed using an Olympus Fluoview BX50WI inverted microscope, and MP-IVM in live anaesthetized mice was performed using an upright microscope (Prairie Technologies) with a MaiTai Ti:Sapphire laser (Spectra-Physics). All animal studies were approved by the Institutional Animal Care and Use Committee of Harvard Medical School and complied with National Institutes of Health guidelines.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 27 August 2013; accepted 4 March 2014.**

**Published online 23 April; corrected online 4 June 2014 (see full-text HTML version for details).**

- Nestle, F. O., Di Meglio, P., Qin, J. Z. & Nickoloff, B. J. Skin immune sentinels in health and disease. *Nature Rev. Immunol.* **9**, 679–691 (2009).
- Perera, G. K., Di Meglio, P. & Nestle, F. O. Psoriasis. *Annu. Rev. Pathol.* **7**, 385–422 (2012).
- Cai, Y. *et al.* Pivotal role of dermal IL-17-producing γδ T cells in skin inflammation. *Immunity* **35**, 596–610 (2011).
- Gray, E. E., Suzuki, K. & Cyster, J. G. Cutting edge: identification of a motile IL-17-producing γδ T cell population in the dermis. *J. Immunol.* **186**, 6091–6095 (2011).
- Rosas-Ballina, M. *et al.* Acetylcholine-synthesizing T cells relay neural signals in a vagus nerve circuit. *Science* **334**, 98–101 (2011).
- Chiu, I. M., von Hehn, C. A. & Woolf, C. J. Neurogenic inflammation and the peripheral nervous system in host defense and immunopathology. *Nature Neurosci.* **15**, 1063–1067 (2012).
- van der Fits, L. *et al.* Imiquimod-induced psoriasis-like skin inflammation in mice is mediated via the IL-23/IL-17 axis. *J. Immunol.* **182**, 5836–5845 (2009).
- Flutter, B. & Nestle, F. O. TLRs to cytokines: mechanistic insights from the imiquimod mouse model of psoriasis. *Eur. J. Immunol.* **43**, 3138–3146 (2013).
- Abrahamsen, B. *et al.* The cell and molecular basis of mechanical, cold, and inflammatory pain. *Science* **321**, 702–705 (2008).
- Lumpkin, E. A. & Caterina, M. J. Mechanisms of sensory transduction in the skin. *Nature* **445**, 858–865 (2007).
- Sándor, K., Helyes, Z., Elekes, K. & Szolcsanyi, J. Involvement of capsaicin-sensitive afferents and the transient receptor potential vanilloid 1 receptor in xylene-induced nocifensive behaviour and inflammation in the mouse. *Neurosci. Lett.* **451**, 204–207 (2009).
- Zheng, Y. *et al.* Interleukin-22, a T<sub>H</sub>17 cytokine, mediates IL-23-induced dermal inflammation and acanthosis. *Nature* **445**, 648–651 (2007).
- Lowes, M. A., Russell, C. B., Martin, D. A., Towne, J. E. & Krueger, J. G. The IL-23/T17 pathogenic axis in psoriasis is amplified by keratinocyte responses. *Trends Immunol.* **34**, 174–181 (2013).
- Rapp, S. R., Feldman, S. R., Exum, M. L., Fleischer, A. B. Jr & Reboussin, D. M. Psoriasis causes as much disability as other major medical diseases. *J. Am. Acad. Dermatol.* **41**, 401–407 (1999).
- Farber, E. M., Lanigan, S. W. & Boer, J. The role of cutaneous sensory nerves in the maintenance of psoriasis. *Int. J. Dermatol.* **29**, 418–420 (1990).
- Ostrowski, S. M., Belkadi, A., Loyd, C. M., Diaconu, D. & Ward, N. L. Cutaneous denervation of psoriasiform mouse skin improves acanthosis and inflammation in a sensory neuropeptide-dependent manner. *J. Invest. Dermatol.* **131**, 1530–1538 (2011).
- Grebe, K. M. *et al.* Sympathetic nervous system control of anti-influenza CD8<sup>+</sup> T cell responses. *Proc. Natl Acad. Sci. USA* **106**, 5300–5305 (2009).
- Swirski, F. K. *et al.* Identification of splenic reservoir monocytes and their deployment to inflammatory sites. *Science* **325**, 612–616 (2009).
- Weninger, W. *et al.* Specialized contributions by α(1,3)-fucosyltransferase-IV and FucT-VII during leukocyte rolling in dermal microvessels. *Immunity* **12**, 665–676 (2000).
- Iannaccone, M. *et al.* Subcapsular sinus macrophages prevent CNS invasion on peripheral infection with a neurotropic virus. *Nature* **465**, 1079–1083 (2010).
- Matloubian, M. *et al.* Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature* **427**, 355–360 (2004).
- Gray, E. E. *et al.* Deficiency in IL-17-committed V<sub>γ</sub>4<sup>+</sup> γδ T cells in a spontaneous Sox13-mutant CD45.1<sup>+</sup> congenic mouse substrain provides protection from dermatitis. *Nature Immunol.* **14**, 584–592 (2013).
- Pantelyushin, S. *et al.* Rorγt<sup>+</sup> innate lymphocytes and γδ T cells initiate psoriasiform plaque formation in mice. *J. Clin. Invest.* **122**, 2252–2256 (2012).
- Guo, L., Junttila, I. S. & Paul, W. E. Cytokine-induced cytokine production by conventional and innate lymphoid cells. *Trends Immunol.* **33**, 598–606 (2012).
- Awasthi, A. *et al.* Cutting edge: IL-23 receptor GFP reporter mice reveal distinct populations of IL-17-producing cells. *J. Immunol.* **182**, 5904–5908 (2009).
- Muller, G. *et al.* IL-12 as mediator and adjuvant for the induction of contact sensitivity *in vivo*. *J. Immunol.* **155**, 4661–4668 (1995).
- Zigmond, E. *et al.* Ly6C<sup>hi</sup> monocytes in the inflamed colon give rise to proinflammatory effector cells and migratory antigen-presenting cells. *Immunity* **37**, 1076–1090 (2012).
- Guilliams, M. *et al.* From skin dendritic cells to a simplified classification of human and mouse dendritic cell subsets. *Eur. J. Immunol.* **40**, 2089–2094 (2010).
- Wohn, C. *et al.* Langerin<sup>neg</sup> conventional dendritic cells produce IL-23 to drive psoriatic plaque formation in mice. *Proc. Natl Acad. Sci. USA* **110**, 10723–10728 (2013).
- Chiu, I. M. *et al.* Bacteria activate sensory neurons that modulate pain and inflammation. *Nature* (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank G. Cheng, M. Flynn and S. Omid for technical support; M. Perdue, L. Jones and E. Nigro for secretarial assistance; E. Gray for providing detailed protocols for flow cytometry of skin samples; J. Harris for discussion and providing detailed RNA isolation protocols of skin samples; R. T. Roderick for histopathology analysis; T. Liu and J. Ru-Rong for providing the RTX denervation protocol; M. Ghebremichael for statistical advice; F. Winau and J. L. Rodriguez-Fernandez for critical reading of the manuscript and members of the von Andrian laboratory for discussion and advice. This work was supported by National Institutes of Health (NIH) grants AI069259, AI078897, AI095261 and AI111595 (to U.H.v.A.), NIH 5F31AR063546-02 (to J.O.-M.) and the Human Frontiers Science Program, Charles A. King Trust and National Psoriasis Foundation (to L.R.-B.).

**Author Contributions** L.R.-B., J.O.-M., and U.H.v.A. designed the study. L.R.-B., J.O.-M., M.P., E.N., A.T. and D.A. performed and collected data from experiments, and L.R.-B., J.O.-M., A.T. and E.N. analysed data. J.N.W. provided reagents and gave conceptual advice. S.P. contributed to experiments in Extended Data Fig. 7d and provided technical support. J.O.-M., L.R.-B. and U.H.v.A. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to U.H.v.A. ([uva@hms.harvard.edu](mailto:uva@hms.harvard.edu)).

## METHODS

**Mice.** C57BL/6 mice, 4–8 weeks old, were purchased from Charles River or the Jackson Laboratory and female mice were used in experiments. IL-23R<sup>GFP/GFP</sup> mice<sup>25</sup>, in which GFP is knocked into the cytoplasmic tail of IL-23R and the homozygous GFP mouse acts as a functional receptor knockout, were provided by M. Oukka and both male and female mice were used in experiments. CD11c-YFP mice<sup>31</sup> were a gift from M. Nussenzweig and both male and female mice were used in experiments. LT $\alpha^{-/-}$  mice<sup>32</sup> were purchased from the Jackson Laboratory and male mice were used in experiments. CD11c-DTR<sup>33</sup> mice were purchased from Jackson Laboratory and both male and female mice were used in experiments. Na<sub>v</sub>1.8-Cre mice were previously described<sup>9</sup>. Rosa26-DTA mice and Rosa26-TdT mice, which express a floxed-STOP cassette upstream of the ubiquitously expressed Rosa26-DTA or Rosa26-TdT construct, respectively, were purchased from Jackson Laboratory. Na<sub>v</sub>1.8-Cre male mice were bred with Rosa26-DTA and Rosa26-TdT female mice to generate Na<sub>v</sub>1.8-DTA for functional studies and Na<sub>v</sub>1.8-TdT for imaging experiments, respectively, of which both matched male and female litters were used in experiments. Glial fibrillary acidic protein (GFAP)-TdT mice were generated by crossing heterozygous GFAP-Cre mice<sup>34</sup> purchased from Jackson Laboratory with Rosa26-TdT mice. MHC-II-GFP mice<sup>35</sup> were provided by M. Boes. Bone marrow chimaeras were generated by irradiating Na<sub>v</sub>1.8-TdT mice or GFAP-TdT mice with two split doses totalling 1,300 rad and reconstituting with CD11c-YFP or MHC-II-GFP unfractionated bone marrow injected intravenously, respectively. Bone marrow chimaeras were allowed to rest for 12 weeks before use. Mice were all housed in specific pathogen-free conditions in accordance with the National Institutes of Health and all experimental animal protocols were approved by the Institutional Animal Care and Use Committee at Harvard Medical School. For most animal experiments, because the contralateral ear served as a control for the ear in which an inflammatory stimulus was applied, no randomization was used. Investigators were blinded for the initial ear-swelling experiments and subsequently no blinding was used.

**Denervation.** RTX, a capsaicin analogue, was injected subcutaneously into the flank of 4-week-old mice in three escalating (30  $\mu\text{g kg}^{-1}$ , 70  $\mu\text{g kg}^{-1}$  and 100  $\mu\text{g kg}^{-1}$ ) doses on consecutive days<sup>11</sup>. Control mice were treated with vehicle solution (DMSO in PBS). Mice were allowed to rest for 4 weeks before denervation was confirmed by tail-flick assay. The tail-flick assay was conducted by holding mice vertically in a relaxed fashion allowing their tail to be immersed in a temperature-controlled water bath maintained at 52 °C. Denervated mice exhibited a tail-flick latency of >10 s. Besides insensitivity to noxious heat stimuli, overall behaviour qualitatively remained unaltered in RTX-treated mice. For chemical sympathectomy, mice were intraperitoneally injected with 80  $\text{mg kg}^{-1}$  6-OHDA in 0.01% ascorbic acid in PBS at day -3. Control mice received injections of 0.01% ascorbic acid in PBS<sup>17</sup>.

**IMQ and DNFB treatment, IL-23 injection and ear measurement.** Mice, 8–12 weeks old, of indicated genotypes or pharmacological treatments, were treated with 25 mg of 5% IMQ applied topically to dorsal and ventral aspects of ear skin totalling 1,250  $\mu\text{g}$  of IMQ per day. Mice were treated starting on day 0 three times and killed for analyses on day 3 or were treated six times and killed for analyses on day 6. IL-23 was injected intradermally into the dorsal aspect of the ear skin as a 50  $\mu\text{g ml}^{-1}$  solution in 10  $\mu\text{l}$  PBS (500 ng per ear). Contralateral ears were injected with PBS alone. A 0.5% solution of DNFB in acetone was applied to the dorsal and ventral aspects of ear skin without prior sensitization. Ear thickness was measured using an engineer's micrometer (Mitutoyo) at the indicated time points and the change in ear thickness is calculated as the change in ear thickness from the treated ear relative to the contralateral naive or vehicle-treated ear.

**FTY720 treatment and depletion of myeloid cells.** FTY720 (1  $\text{mg kg}^{-1}$ ), an agonist which results in S1P receptor internalization, or PBS was injected daily intraperitoneally starting from day 0 of IMQ challenge. Efficacy of treatment was confirmed by a marked reduction in circulating lymphocytes from peripheral blood (data not shown). To deplete dendritic cells from skin, CD11c-DTR mice<sup>33</sup> were injected at day -1 with 4 ng of DTX per g mouse, a dose that is nontoxic to murine cells not expressing the DTR<sup>36</sup>. Depletion of DDCs and Langerhans cells, but not macrophages, was confirmed by flow cytometry (Extended Data Fig. 8). To deplete neutrophils and inflammatory monocytes, 500  $\mu\text{g}$  of anti-Gr-1 per mouse (clone RB6-8C5; BioXCell), was injected intraperitoneally at days -1, 0, 1 and 2. Depletion was confirmed by flow cytometry of peripheral blood as well as challenged ear skin showing a paucity of neutrophils and inflammatory monocytes (Extended Data Fig. 8).

**Histology.** Ears were embedded in paraffin and submitted for histological analysis by haematoxylin and eosin staining to the Harvard Rodent Histopathology Core.

**Whole-mount immunofluorescence of ear skin.** Ears were harvested and split into dorsal and ventral halves. After fixation in 4% paraformaldehyde, any adherent cartilage was removed under a dissecting microscope to expose the dermis evenly for imaging. Tissue was blocked in blocking buffer containing PBS with 0.5% BSA, 0.3% Triton X-100, 10% goat serum and Fc Block and also stained with the following antibodies in the same buffer. Unconjugated antibodies used include: anti-neuronal

class III  $\beta$ -tubulin (clone TUJ1; Covance), anti-TH (clone A2B5-105; Millipore), anti-Lyve-1 (clone ALY7; eBiosciences), anti-peripherin (Polyclonal; Abcam) and anti-NeuN (A60; Millipore). Alexa488-conjugated antibodies include: anti-neuronal class III  $\beta$ -tubulin (clone TUJ1; Covance), goat anti-mouse IgG (Invitrogen) and goat anti-rabbit IgG. Alexa647-conjugated antibodies include: goat anti-rat IgG and goat anti-rabbit IgG. Washing steps were performed in PBS with 0.2% BSA, and 0.1% Triton X-100. Ears were mounted with the dermis facing the imaging plane in FluorSave reagent (Calbiochem).

**Confocal microscopy.** Confocal images were acquired on an Olympus Fluoview BX50WI inverted microscope with  $\times 10/0.4$ ,  $\times 20/0.5$  and  $\times 40/1.3$  magnification/numerical aperture objectives. For images used in three-dimensional analysis, image planes were acquired at 0.5  $\mu\text{m}$  intervals through the imaging volume.

**Intravital two-photon microscopy.** Anaesthetized mice were placed on a custom-built stage and the ear was fixed to a temperature-controlled metallic support to facilitate exposure of the dorsal aspect to the water-immersion  $\times 20$  objective (0.95 numerical aperture) of an upright microscope (Prairie Technologies). A MaiTai Ti:sapphire laser (Spectra-Physics) was tuned between 870 nm and 900 nm for multiphoton excitation and second-harmonic generation. For dynamic analysis of cell interaction in four dimensions, several  $x$ - $y$  sections ( $512 \times 512$ ) with  $z$  spacing ranging from 2  $\mu\text{m}$  to 4  $\mu\text{m}$  were acquired every 15–20 s with an electronic zoom varying from  $1 \times$  to  $3 \times$ . Emitted light and second-harmonic signals were directed through 450/80 nm, 525/50 nm and 630/120 nm bandpass filters and detected with non-descanned detectors. Post-acquisition image analysis, volume rendering and four-dimensional time-lapse videos were performed using Imaris software (Bitplane scientific software).

**Intravital microscopy and image analysis.** Intravital microscopy of skin was performed as previously described<sup>19</sup>. In brief, control or RTX-treated mice were anaesthetized and the left ear was exposed and positioned for epifluorescence intravital microscopy. Preparations were transferred to an intravital microscope (IV-500; Mikron Instruments), equipped with a Rapp OptoElectronic SP-20 xenon flash lamp system and QImaging Rolera-MGi EMCCD camera. The fluorescent dye rhodamine-6G (20  $\text{mg kg}^{-1}$  in PBS) was administered through the catheterized right carotid artery to visualize circulating leukocytes. Cell behaviour in skin venules was recorded in 10 min recordings through  $\times 10$  or  $\times 20$  water-immersion objectives (Achromplan; Carl Zeiss). Rolling fractions in individual vessel segments were determined offline by playback of digital video files. The rolling fraction was determined as the percentage of cells interacting with skin venules in the total number of cells passing through a vessel during the observation period.

**Cytokine quantification by ELISA.** Skin biopsies from ears were obtained using 10 mm diameter skin biopsy punches (Acuderm). Samples were homogenized in Tissue Extraction Reagent I (Invitrogen) in the presence of protease inhibitor cocktail (Roche) using a gentleMACS dissociator (Miltenyi Biotec). Skin protein extracts were assayed for IL-17A, IL-22, IL-23p40 (Biolegend) and IL-17F (R&D Systems) in accordance with the manufacturer's instructions.

**RNA isolation and qPCR.** RNA from sorted and pelleted cells was isolated using RNeasy Plus Mini Kit (Qiagen) including a gDNA elimination step. Ear skin and dorsal root ganglia (DRGs) were harvested and placed immediately in RNALater (Ambion) before homogenization and RNA isolation using Qiagen RNeasy Plus Mini Kit (Qiagen)<sup>9</sup>. DRGs for comparison of the efficacy of denervation were harvested from equivalent anatomical locations, typically consisting of the cervical and thoracic ganglia from C1–T2. cDNA synthesis was done using Superscript Vilo cDNA synthesis kit (Invitrogen) following the manufacturer's instructions. Relative quantification of transcripts was done using validated Quantitect Primer Assays (Qiagen) combined with the QuantiTect SYBR Green Detection Kit (Qiagen) on a LightCycler 480 (Roche). Relative expression of genes was calculated to *Gapdh* using the  $\Delta\Delta C_T$  method.

**Tissue digestion.** Single-cell suspensions of lymph nodes, spleen and bone marrow were generated as described previously<sup>20</sup>. For ear skin digestion, the dorsal and ventral aspects of the ear were mechanically separated before mincing and placing into a digestion mix modified from that described previously<sup>4</sup>. Ears were digested for 80 min at 37 °C in gentleMACS tubes (Miltenyi) with gentle agitation in freshly prepared digestion mix consisting of DMEM (Gibco) supplemented with HEPES (Invitrogen), 2% FCS, 100  $\mu\text{g ml}^{-1}$  Liberase TM (Roche), 100  $\mu\text{g ml}^{-1}$  DNase I (Roche) and 0.5  $\text{mg ml}^{-1}$  Hyaluronidase (Sigma). After enzymatic digestion, the mixture was processed using a gentleMACS homogenizer (Miltenyi) to obtain a cell suspension, which was then filtered through a 70  $\mu\text{m}$  cell strainer (BD). Cells were then resuspended in FACS buffer for analysis. If cell suspensions were to be analysed for cytokine-producing cells, ears were harvested and digested in the presence of Brefeldin A (Biolegend).

**Flow cytometry, cell sorting and cell counts.** Single-cell suspensions in FACS Buffer (PBS with 2 mM EDTA and 2% FCS (Invitrogen-GIBCO)) were pre-incubated with Fc Block (clone 2.4G2) before staining for surface antigens. FITC-conjugated antibodies used include: anti-Ly-6G (clone 1A8; BD Pharmingen). Alexa488-conjugated



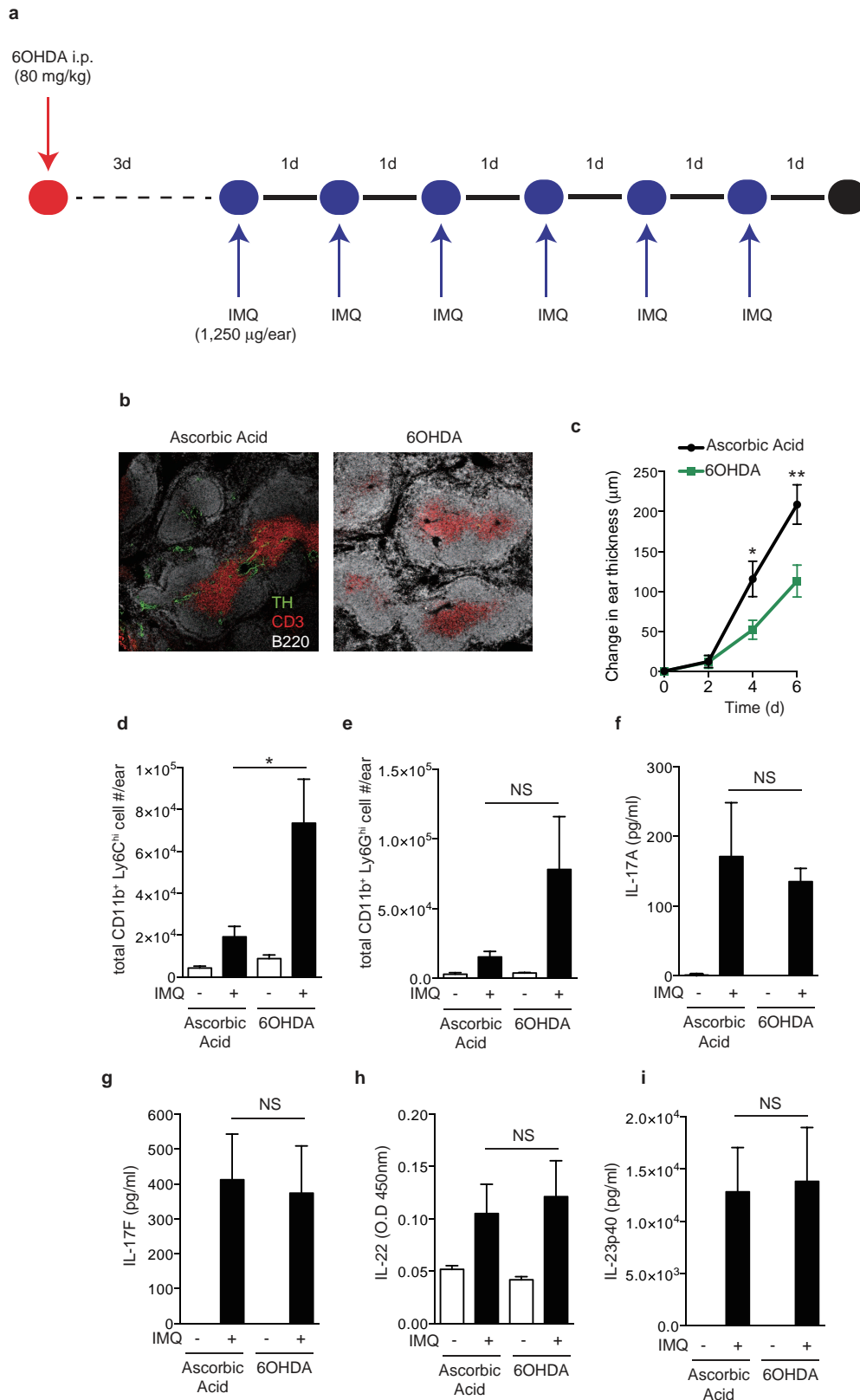
antibodies used include: anti-CD103 (clone 2E7; Biolegend). PE-conjugated antibodies used include: anti- $\delta$ -TCR (clone GL3; Biolegend), anti-Ly-6C (clone HK1.4; Biolegend). PerCP/Cy5.5-conjugated antibodies used include: anti-CD45.2 (clone 104; Biolegend). PE-Cy7-conjugated antibodies include: anti-TCR- $\beta$  (clone H57-597; Biolegend), anti-CD11c (clone HL3; BD Pharmingen). Alexa647-conjugated antibodies used include: anti-CCR6 (clone 140706; BD Pharmingen), anti-CD11b (clone M1/70; Biolegend). APC-Cy7-conjugated antibodies used include: anti-Thy1.2/CD90.2 (clone 30-H12; Biolegend), anti-I-A/I-E 'Class-II' (clone M5/114.15.2; Biolegend). Cells were then washed with PBS and resuspended in MACS buffer for immediate acquisition or fixed in Cytofix (BD Pharmingen) as per the manufacturer's instructions for later acquisition. For analysis, cells were acquired on a BD FACS CANTO (BD Pharmingen) and analysed using FlowJo software (Treestar). For intracellular cytokine staining, cells were not restimulated and rather harvested and digested in the presence of Brefeldin A (Biolegend). After staining for surface antigens, cells were then fixed and permeabilized using BD Cytofix/Cytoperm kit (BD Pharmingen) as per the manufacturer's instructions. Cells were stained in Perm/Wash buffer. Alexa488-conjugated antibodies used include: anti-IL-17A (clone TC11-18H10.1; Biolegend), anti-IL-17F (clone 9D3.1C8; Biolegend). Alexa647-conjugated antibodies used include: anti-IL-17A (clone TC11-18H10.1; Biolegend), anti-IL-17F (clone 9D3.1C8; Biolegend) and anti-IL-22 (clone Poly5164; Biolegend). For determining total counts of cell subsets, an aliquot of the same cell suspension used for flow cytometry was stained for CD45.2 and acquired on an Accuri Cytometer (BD Biosciences) with a known amount of CountBright counting beads (Invitrogen). The total CD45<sup>+</sup> cell number was then determined in the original cell suspension and used for total quantification of the cell number of each subset of interest defined by multi-parameter flow cytometry based on that subset's frequency relative to the CD45<sup>+</sup> population. For cell sorting, cells were stained for indicated surface markers and sorted using a BD FACS Aria (BD Biosciences) into complete DMEM media before RNA extraction.

**Image analysis.** Images from confocal microscopy and from two-photon microscopy were analysed on either Volocity software (Improvision) or Imaris (Bitplane).

For determining the distance of CD11c-YFP cells from nerves, lymphatics or blood vessels, the centroid of dendritic cells was calculated by an unbiased approach, and the distance between a given dendritic cell centroid and the closest nerve, blood vessel or lymphatic vessel was calculated using Imaris software. To bin cells into contact (<0  $\mu$ m), proximal (0–7  $\mu$ m) and distal (>7  $\mu$ m) fractions, the calculated average radius of a dendritic cell (7.04  $\mu$ m) was subtracted from each measured distance.

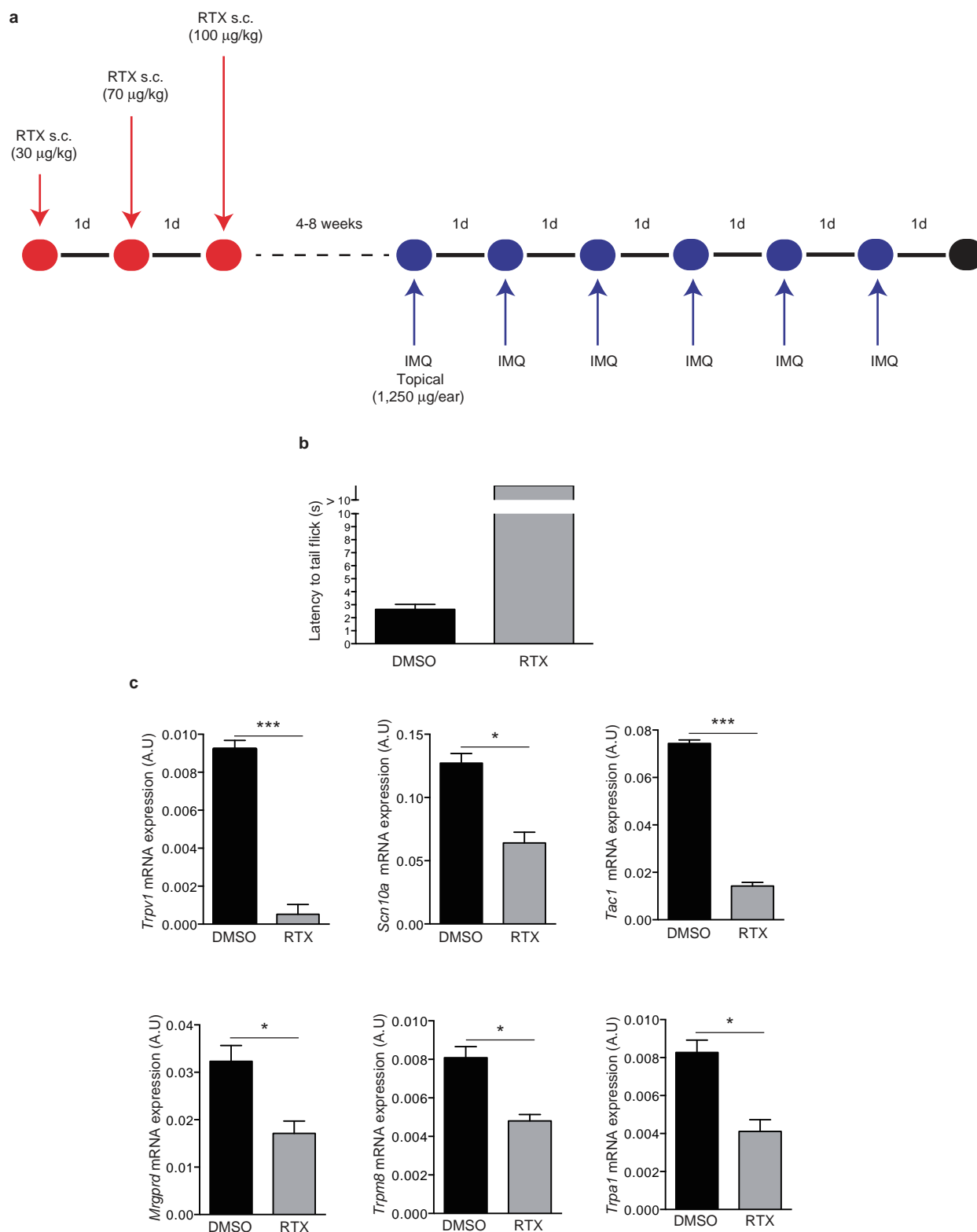
**Statistical analyses.** Precise experimental numbers of animals are reported in the figure legends. Experiments were repeated at least three times except in Figs 2e and 3e–g in which two replicates were done assessing ten mice total in each experimental group overall. Some data, such as ear-swelling curves, represent pooled averages of the sum total of animals used in experiments whereas other data consist of a representative experiment of the independent experiments. All statistical analyses were performed using Prism (GraphPad Software) and results are calculated as means with error bars representing the s.e.m. Means between two groups were compared by using a two-tailed *t*-test. Means between three or more groups were compared by using a one-way or two-way ANOVA. A Chi-squared statistical analysis was performed for Fig. 4d comparing the total number of dendritic cells in contact, proximal and distal bins relative to nerves versus lymphatic vessels or versus blood vessels.

31. Lindquist, R. L. *et al.* Visualizing dendritic cell networks *in vivo*. *Nature Immunol.* **5**, 1243–1250 (2004).
32. De Togni, P. *et al.* Abnormal development of peripheral lymphoid organs in mice deficient in lymphotoxin. *Science* **264**, 703–707 (1994).
33. Jung, S. *et al.* *In vivo* depletion of CD11c<sup>+</sup> dendritic cells abrogates priming of CD8<sup>+</sup> T cells by exogenous cell-associated antigens. *Immunity* **17**, 211–220 (2002).
34. Garcia, A. D., Doan, N. B., Imura, T., Bush, T. G. & Sofroniew, M. V. GFAP-expressing progenitors are the principal source of constitutive neurogenesis in adult mouse forebrain. *Nature Neurosci.* **7**, 1233–1241 (2004).
35. Boes, M. *et al.* T-cell engagement of dendritic cells rapidly rearranges MHC class II transport. *Nature* **418**, 983–988 (2002).
36. Saito, M. *et al.* Diphtheria toxin receptor-mediated conditional and targeted cell ablation in transgenic mice. *Nature Biotechnol.* **19**, 746–750 (2001).



**Extended Data Figure 1 | 6OHDA treatment ablates sympathetic nerve function and reduces ear swelling, but does not ameliorate the inflammatory response to IMQ treatment.** **a**, The experimental protocol was as follows. Mice were injected intraperitoneally with 6OHDA, resulting in a reversible chemical sympathectomy lasting for approximately 2 weeks. After a rest period of 3 days animals were challenged topically on the ear with IMQ. **b**, Representative section of splenic white pulp showing B cells (B220, white), T cells (CD3, red), and TH<sup>+</sup> (green) nerve fibres in vehicle

(ascorbic acid)-treated and sympathectomized (6OHDA) mice. **c–i**, Analysis of the inflammatory response in ears of vehicle (ascorbic acid)-treated and sympathectomized (6OHDA) mice after daily topical IMQ challenge: timecourse of change in ear thickness of IMQ-treated ear relative to the contralateral ear ( $n = 10$ ;  $*P < 0.05$ ;  $**P < 0.01$ ) (**c**) and total number of infiltrating monocytes (**d**) and neutrophils (**e**), and the amount of IL-17A (**f**), IL-17F (**g**), IL-22 (**h**) and IL-23p40 (**i**) in protein extracts of IMQ exposed ears at day 3 ( $*P < 0.05$ ;  $n = 5$ ). NS, not significant.

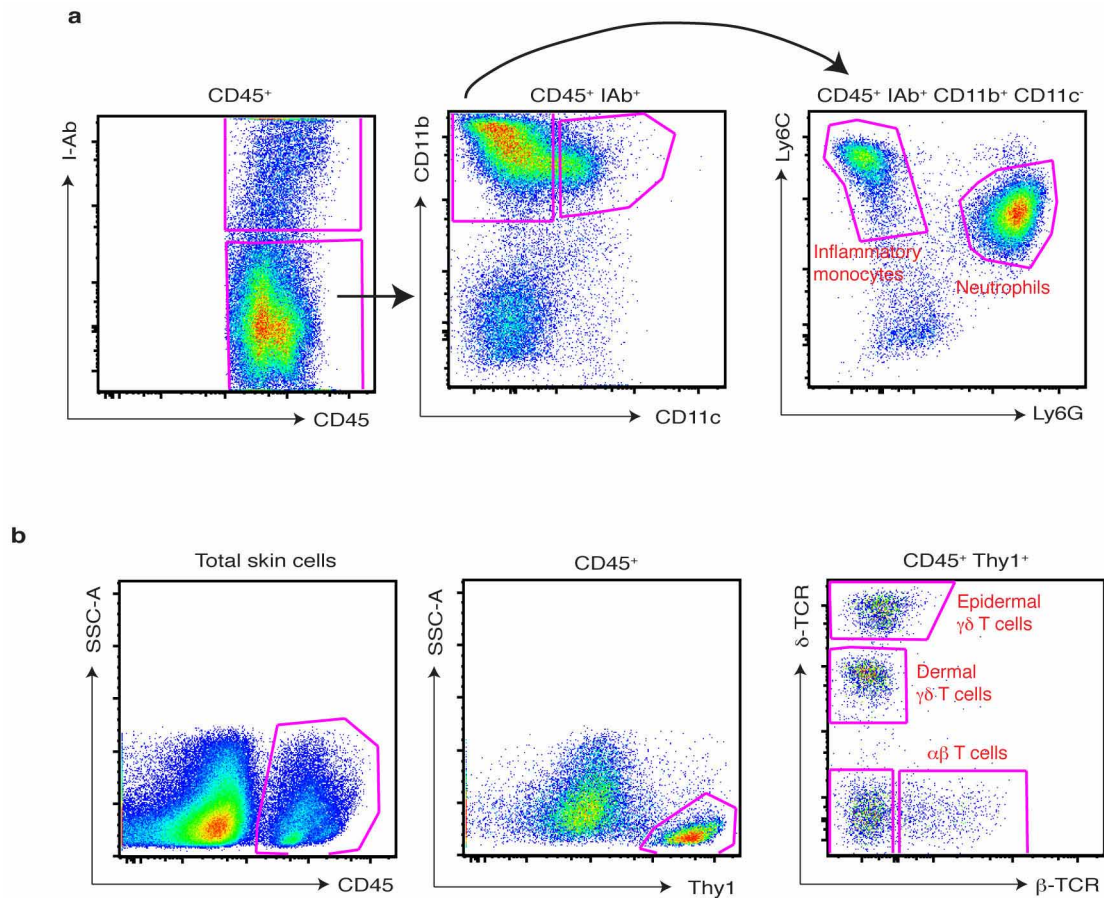


**Extended Data Figure 2 | RTX treatment diminishes noxious heat sensation and decreases the expression of nociceptor markers on dorsal root ganglia.**

**a**, Schematic protocol of nociceptor ablation and induction of psoriasiform skin inflammation. RTX was injected subcutaneously into the back in three escalating doses ( $30 \mu\text{g kg}^{-1}$ ,  $70 \mu\text{g kg}^{-1}$  and  $100 \mu\text{g kg}^{-1}$ ) on consecutive days and mice were allowed to rest for at least 4 weeks before IMQ treatment.

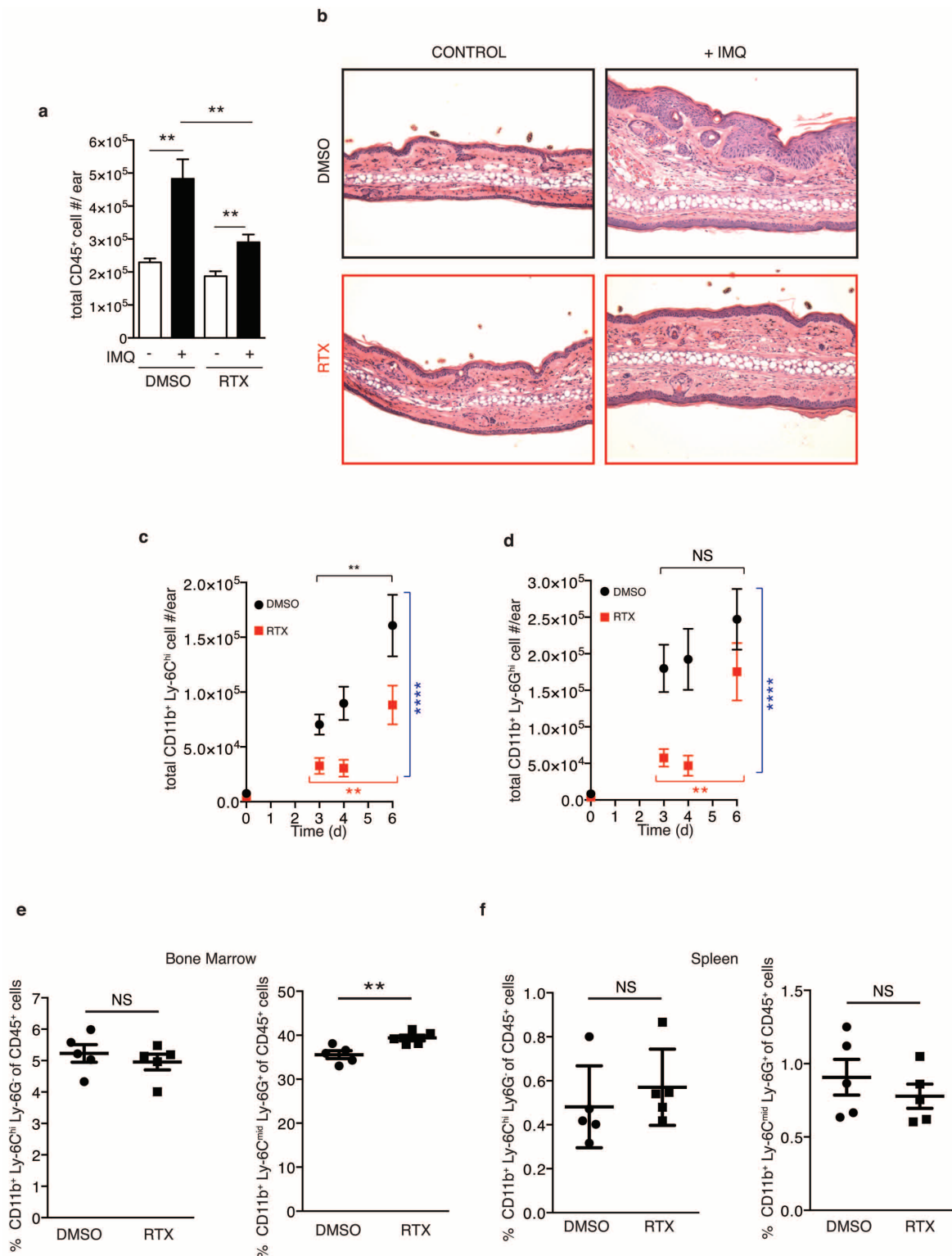
**b**, Denervation was confirmed by immersing the tail of mice into a temperature-controlled water bath maintained at  $52^\circ\text{C}$  and the latency to the first tail movement to avoid water was measured ( $n = 6$ ). **c**, Total RNA was isolated from dorsal root ganglia (level C1–C7) of vehicle (DMSO)- and RTX-treated mice and the levels of *Trpv1*, *Scn10a* ( $\text{Na}_v1.8$ ), *Tac1* (substance P), *Mrgprd*, *Trpm8* and *Trpa1* mRNA relative to *Gapdh* were determined ( $n = 3$ ).





**Extended Data Figure 3 | Gating strategy for myeloid cells and T-cell subsets from digested ear skin.** **a**, The ear skin of mice challenged for 3 days with IMQ was digested as described in Methods and, after doublet exclusion and gating on defined FSC-A, SSC-A parameters, infiltrating myeloid cells were gated as CD45<sup>+</sup> I-Ab (Class-II)<sup>-</sup>, CD11b<sup>+</sup> CD11c<sup>-</sup>, and then subdivided into

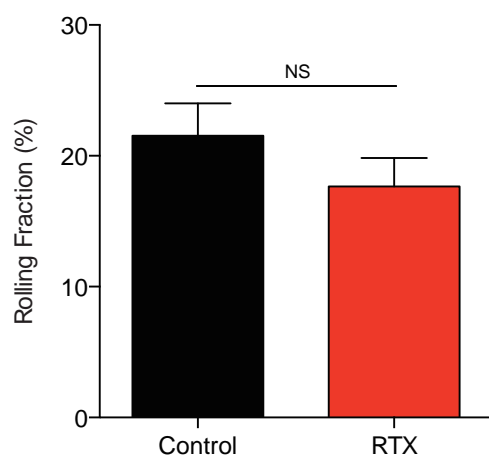
inflammatory monocytes and neutrophils based on Ly-6C and Ly-6G staining. **b**, The ear skin of naive mice was digested as described in Methods and, after doublet exclusion and gating on defined FSC-A, SSC-A parameters, cutaneous T cells were gated on CD45<sup>+</sup>, Thy1<sup>+</sup>, and then divided into subsets based on staining for  $\delta$ -TCR and  $\beta$ -TCR.



**Extended Data Figure 4 | RTX treatment reduces the immune cell infiltrate upon IMQ treatment in the skin but does not affect reservoirs of inflammatory monocytes and neutrophils at steady state.**

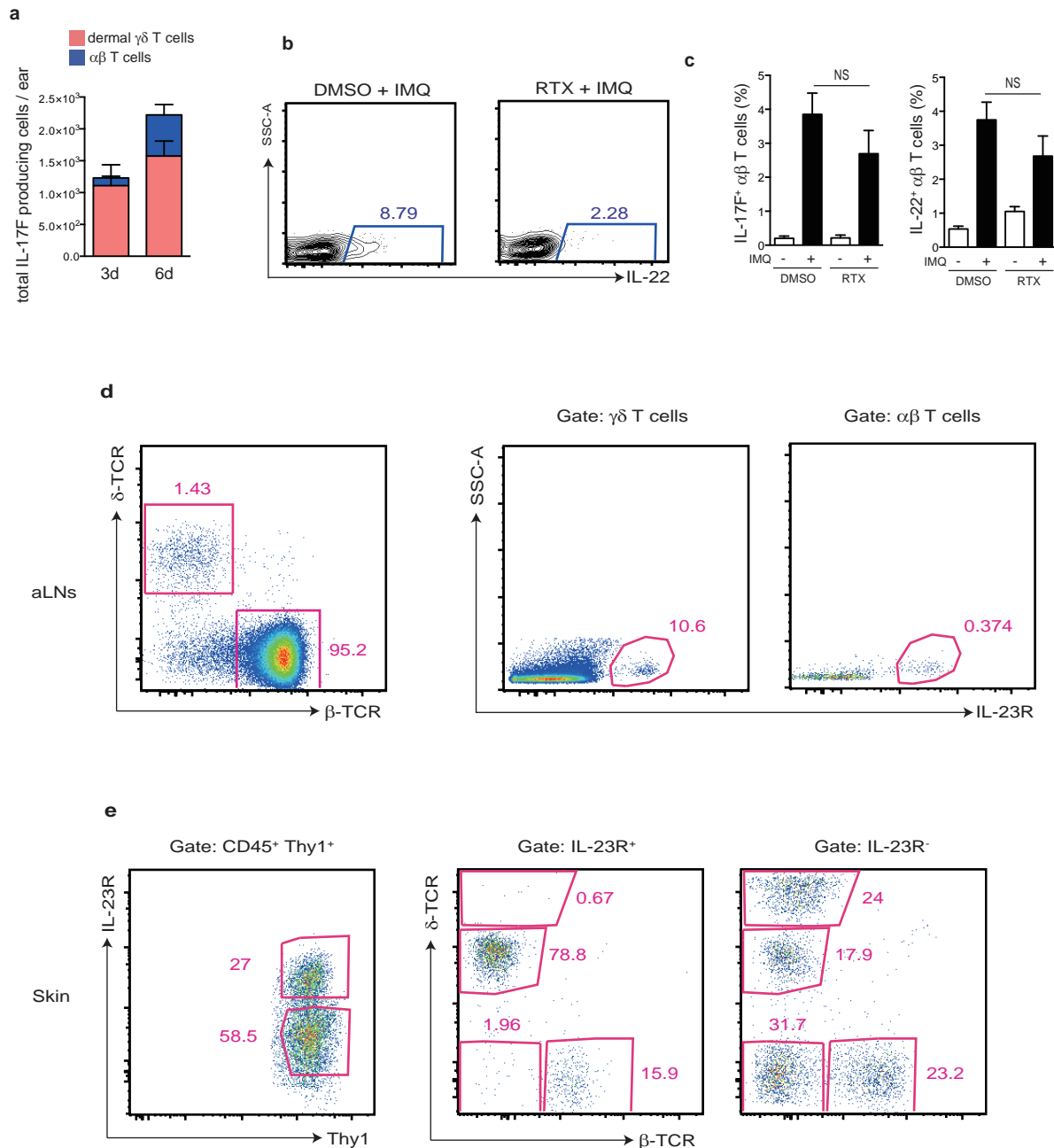
**a**, The ear skin of vehicle (DMSO) or sensory denervated (RTX) mice was treated with topical IMQ cream daily and the total numbers of CD45<sup>+</sup> cells were determined on day 3 as explained in Methods ( $n = 10$ ). **b**, Representative histological sections of untreated and IMQ-treated ears at day 3 stained by haematoxylin and eosin ( $\times 20$ ) ( $n = 5$  per condition). **c**, **d**, Total inflammatory monocytes (CD45<sup>+</sup>, CD11b<sup>+</sup>, Ly-6C<sup>high</sup>) and neutrophils (CD45<sup>+</sup>, CD11b<sup>+</sup>, Ly-6C<sup>mid</sup>) were determined by flow cytometry ( $n = 5$ –10 mice per time point). Two-way ANOVA was run to compare total numbers of inflammatory monocytes and

neutrophils between DMSO and RTX conditions over days 3–6 (\*\*\*\* $P < 0.0001$ ). One-way ANOVA was run to compare total inflammatory monocytes and neutrophil numbers over days 3–6 within DMSO or RTX conditions (\*\* $P < 0.003$ ). **e**, Bone marrow was isolated from wild-type and RTX mice from one femur and the frequency of inflammatory monocytes (CD45<sup>+</sup>, CD11b<sup>+</sup>, Ly-6C<sup>high</sup>, Ly-6C<sup>mid</sup>) and neutrophils (CD45<sup>+</sup>, CD11b<sup>+</sup>, Ly-6C<sup>mid</sup>, Ly-6C<sup>high</sup>) relative to CD45<sup>+</sup> cells was determined by flow cytometry ( $n = 5$ ). **f**, Spleens from wild-type and RTX mice were processed for flow cytometry and the frequency of inflammatory monocytes (CD45<sup>+</sup>, CD11b<sup>+</sup>, Ly-6C<sup>high</sup>, Ly-6C<sup>mid</sup>) and neutrophils (CD45<sup>+</sup>, CD11b<sup>+</sup>, Ly-6C<sup>mid</sup>, Ly-6C<sup>high</sup>) relative to CD45<sup>+</sup> cells was determined ( $n = 5$ ). NS, not significant.



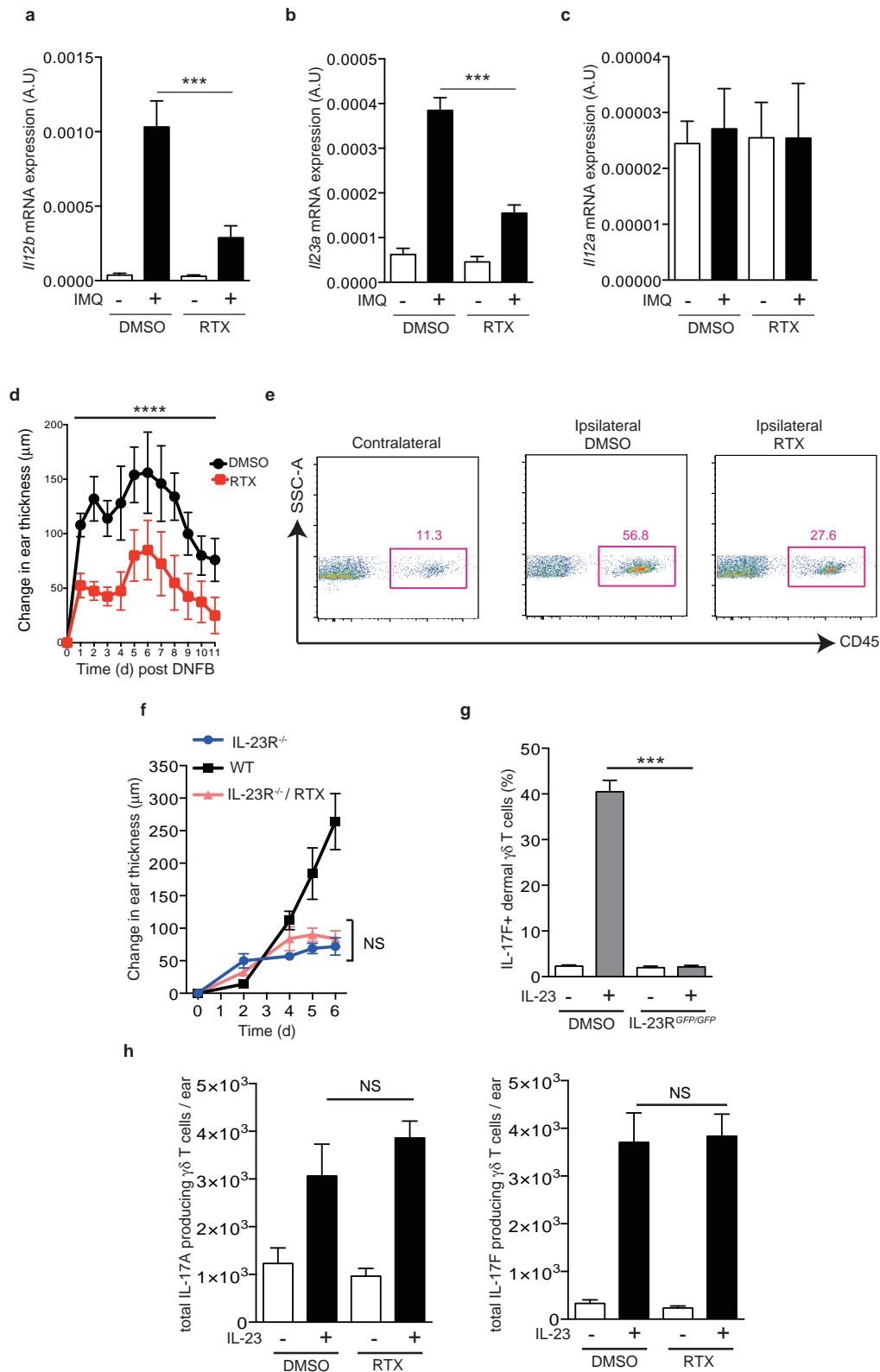
**Extended Data Figure 5 | Leukocyte rolling fractions in skin venules of control and RTX-treated mice analysed by intravital microscopy.** Combined results are shown for 26 venules from 5 control mice and for 20 venules from 4 RTX-treated mice. Data are expressed as mean  $\pm$  s.e.m. of four experiments. NS, not significant.





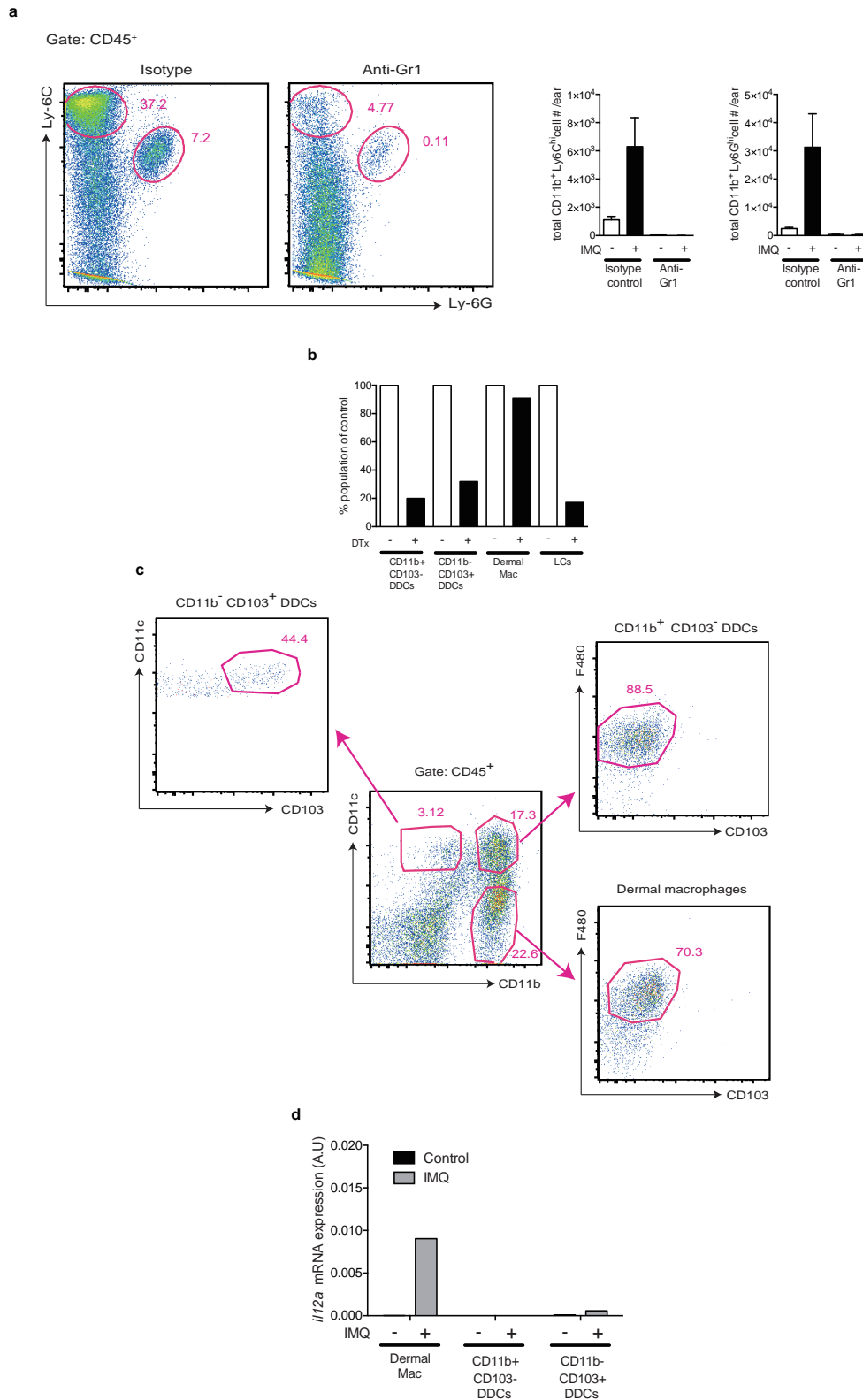
**Extended Data Figure 6 | Dermal  $\gamma\delta$  T cells represent a major source of IL-17F and IL-22 in skin during IMQ challenge and already express IL-23R at steady state.** **a**, Wild-type mice were challenged with IMQ and the total numbers of IL17F<sup>+</sup> dermal  $\gamma\delta$  T cells and  $\alpha\beta$  T cells at 3 days ( $n = 15$ ) or 6 days ( $n = 10$ ) were quantified. **b**, Representative flow plots (related to those depicted in Fig. 2g) of gating for IL-22<sup>+</sup> cells within dermal  $\gamma\delta$  T cells after 6 days of IMQ treatment. **c**, Ears of DMSO- or RTX-treated mice were exposed for 6 days to IMQ and the frequency of IL-17F<sup>+</sup> and IL-22<sup>+</sup> cells within  $\alpha\beta$  T cells was

determined ( $n = 5$ ). **d**, Auricular lymph node (aLN) cells from IL-23R<sup>GFP/+</sup> mice were analysed by flow cytometry for expression of IL-23R-GFP<sup>+</sup> cells within the  $\gamma\delta$  T cells and  $\alpha\beta$  T cell compartment at steady state (representative FACS plot from eight mice analysed). **e**, The ear skin from IL-23R<sup>GFP/+</sup> mice was digested and analysed by flow cytometry and the distribution of T-cells subsets within IL-23R-GFP<sup>+</sup> and IL-23R-GFP<sup>-</sup> fractions of Thy1<sup>+</sup> cells determined (representative FACS plot from eight mice analysed). NS, not significant.



**Extended Data Figure 7 | TRPV1<sup>+</sup> nociceptors regulate IMQ- and DNFB-induced dermatitis and are upstream of IL-23.** **a–c**, After 3 days of IMQ challenge, ears were harvested and processed for total RNA isolation and *Il12b* (**a**), *Il23a* (**b**) and *Il12a* (**c**) mRNA levels were analysed by qPCR ( $n = 5$ ). **d**, DNFB (0.5% in acetone) was applied topically to DMSO and RTX mice. Time course of change in ear thickness of IMQ-treated ear relative to the contralateral ear is represented ( $n = 10$ ). Two-way ANOVA was run to compare ear swelling under DMSO and RTX conditions over time (\*\*\*\* $P < 0.0001$ ). **e**, Representative FACS plots from ears harvested after 24 h

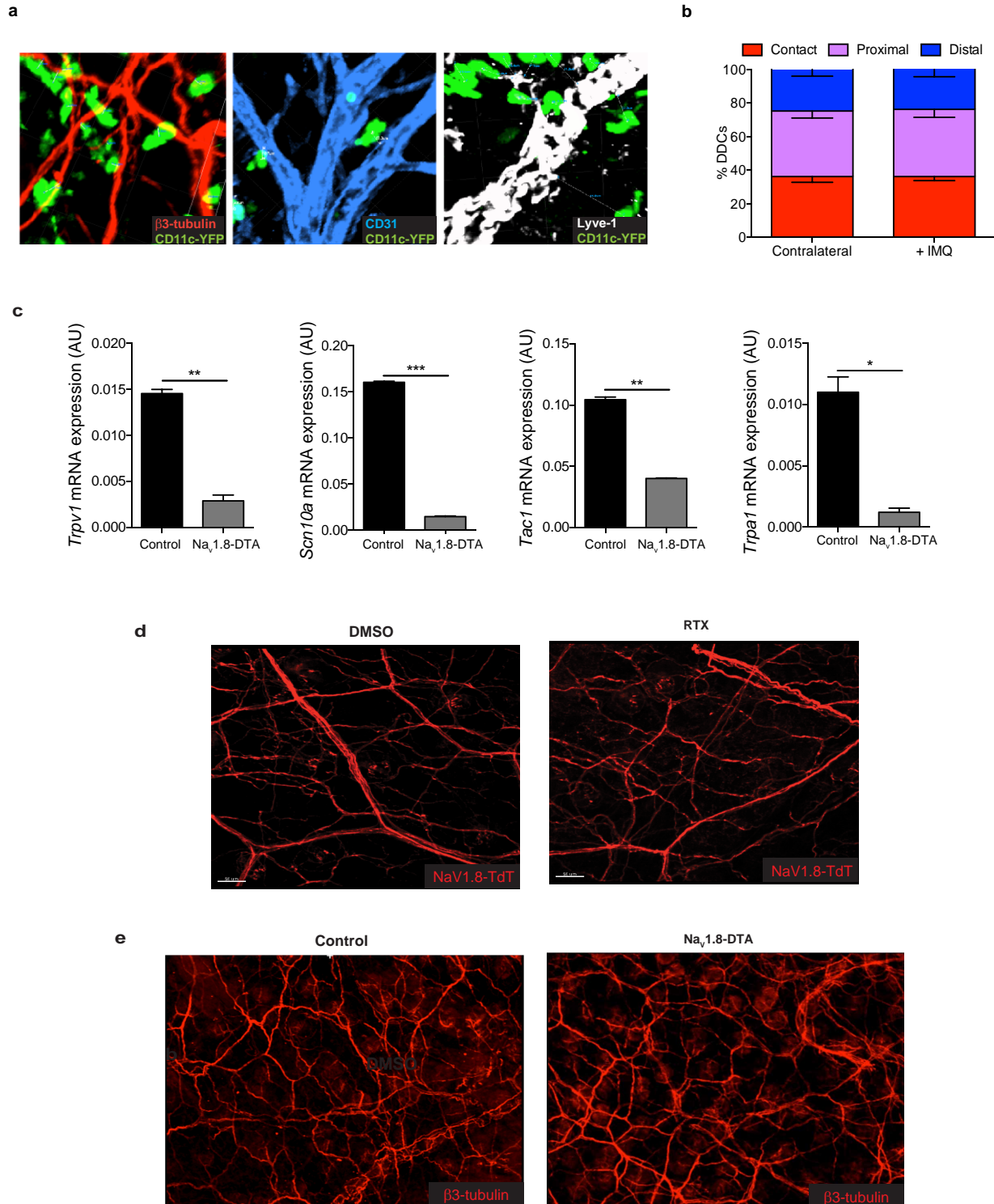
of DNFB application. **f**, IL-23R<sup>GFP/GFP</sup> mice were treated with RTX and then compared to wild-type and IL-23R<sup>GFP/GFP</sup> littermate controls during IMQ treatment. Ear thickness was calculated relative to the contralateral ear ( $n = 5$ ). **g**, After two IL-23 injections into the ear skin of wild-type and IL-23R<sup>GFP/GFP</sup> mice, the frequency of IL-17F<sup>+</sup> cells within dermal  $\gamma\delta$  T cells was determined by flow cytometry ( $n = 5$ ). **h**, IL-23 was injected twice into the ear skin of vehicle- and RTX-treated mice and the total numbers of IL17A<sup>+</sup> or IL-17F<sup>+</sup> dermal  $\gamma\delta$  T cells per ear were determined by flow cytometry ( $n = 5$ ). NS, not significant.



**Extended Data Figure 8 | Selective depletion of migratory and skin-resident myeloid cell subsets in ear skin and gating strategy used for sorting to isolate RNA from MHC-II<sup>+</sup> cells in skin.** **a**, Wild-type mice were treated with anti-Gr-1 (clone RB6-8C5 to deplete neutrophils and inflammatory monocytes) or matched isotype control, challenged with IMQ for 3 days and skin was digested to quantify the total numbers of inflammatory monocytes and neutrophils per ear. Shown are representative plots pre-gated on CD45<sup>+</sup> cells and quantification of cell numbers from  $n = 3$  mice. **b**, DTX treatment resulted in depletion of both subsets of DDCs as well as Langerhans cells (LCs)

but not macrophages. Cells were gated as shown in Extended Data Fig. 8c and normalized to levels in wild-type mice based on the frequency within the CD45<sup>+</sup> population from  $n = 4$  mice. **c**, Ear skin from naive mice was digested and analysed by flow cytometry for the indicated subsets. Shown is a representative plot pre-gated on CD45<sup>+</sup> Class II<sup>+</sup> cells from which further subsets were divided based on CD11b and CD11c expression and then F4/80 and CD103 as indicated. **d**, Total RNA from sorted cells was isolated and qPCR for *Il12a* relative to *Gapdh* was performed from naive and IMQ-treated ears after 6 h from  $n = 20$  pooled mice.

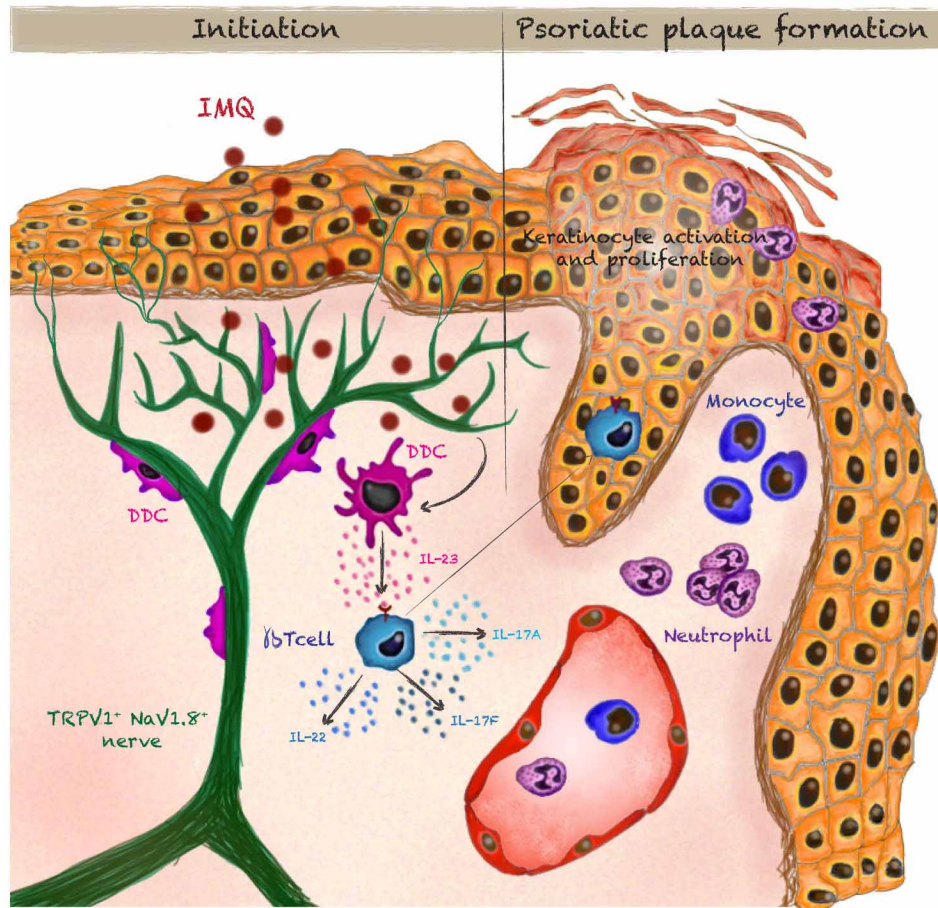




#### Extended Data Figure 9 | DDCs are found in close apposition to Na<sub>v</sub>1.8<sup>+</sup> nociceptors and characterization of RTX-treated and Na<sub>v</sub>1.8-DTA mice.

**a**, Representative confocal micrographs of CD11c-YFP mice stained for β3-tubulin, Lyve-1 (lymphatics) and CD31 (blood and lymphatic endothelial cells). **b**, Three-dimensional quantification of DDC proximity to peripheral nerves in naive and 6 h post-IMQ treatment ears binned into contact (<0 μm), proximal (0–7 μm) and distal (>7 μm) fractions as explained in Methods (*n* of dendritic cells = 200). **c**, Total RNA from dorsal root ganglia (C1–C4) of littermate control and Na<sub>v</sub>1.8-DTA mice was isolated and levels of mRNA for *Trpv1* (TRPV1), *Scn10a* (Na<sub>v</sub>1.8), *Tac1* (substance P) and *Trpa1* (TRPA1) were

determined relative to *Gapdh*. This demonstrates the efficacy of the Na<sub>v</sub>1.8-DTA system and combined with the original reference characterizing the pain phenotype of these mice illustrates that a subset of peptidergic TRPV1<sup>+</sup> nerve fibres is spared. **d**, Representative confocal micrograph of whole-mount ear skin of vehicle- and RTX-treated mice showing preserved nerve scaffold. **e**, Representative confocal micrographs of whole-mount ear skin of control and Na<sub>v</sub>1.8-DTA mice showing preserved nerve scaffold. Although dorsal root ganglia showed a loss of the hallmark ion channels of these nerve subsets (Extended Data Figs 1c, 9c), surprisingly we still observed that RTX mice and Na<sub>v</sub>1.8-DTA mice maintain a meshwork of nerves in the skin.



Extended Data Figure 10 | Summary diagram.

# Ubiquitin is phosphorylated by PINK1 to activate parkin

Fumika Koyano<sup>1,2</sup>, Kei Okatsu<sup>1,2</sup>, Hidetaka Kosako<sup>3</sup>, Yasushi Tamura<sup>4</sup>, Etsu Go<sup>1</sup>, Mayumi Kimura<sup>1</sup>, Yoko Kimura<sup>1,5</sup>, Hikaru Tsuchiya<sup>1</sup>, Hidehito Yoshihara<sup>1</sup>, Takatsugu Hirokawa<sup>6</sup>, Toshiya Endo<sup>7,8</sup>, Edward A. Fon<sup>9</sup>, Jean-François Trempe<sup>10</sup>, Yasushi Saeki<sup>1</sup>, Keiji Tanaka<sup>1</sup> & Noriyuki Matsuda<sup>1,11</sup>

**PINK1** (PTEN induced putative kinase 1) and **PARKIN** (also known as **PARK2**) have been identified as the causal genes responsible for hereditary recessive early-onset Parkinsonism<sup>1,2</sup>. PINK1 is a Ser/Thr kinase that specifically accumulates on depolarized mitochondria, whereas parkin is an E3 ubiquitin ligase that catalyses ubiquitin transfer to mitochondrial substrates<sup>3–5</sup>. PINK1 acts as an upstream factor for parkin<sup>6,7</sup> and is essential both for the activation of latent E3 parkin activity<sup>8</sup> and for recruiting parkin onto depolarized mitochondria<sup>8–12</sup>. Recently, mechanistic insights into mitochondrial quality control mediated by PINK1 and parkin have been revealed<sup>3–5</sup>, and PINK1-dependent phosphorylation of parkin has been reported<sup>13–15</sup>. However, the requirement of PINK1 for parkin activation was not bypassed by phosphomimetic parkin mutation<sup>15</sup>, and how PINK1 accelerates the E3 activity of parkin on damaged mitochondria is still obscure. Here we report that ubiquitin is the genuine substrate of PINK1. PINK1 phosphorylated ubiquitin at Ser 65 both *in vitro* and in cells, and a Ser 65 phosphopeptide derived from endogenous ubiquitin was only detected in cells in the presence of PINK1 and following a decrease in mitochondrial membrane potential. Unexpectedly, phosphomimetic ubiquitin bypassed PINK1-dependent activation of a phosphomimetic parkin mutant in cells. Furthermore, phosphomimetic ubiquitin accelerates discharge of the thioester conjugate formed by UBCH7 (also known as UBE2L3) and ubiquitin (UBCH7~ubiquitin) in the presence of parkin *in vitro*, indicating that it acts allosterically. The phosphorylation-dependent interaction between ubiquitin and parkin suggests that phosphorylated ubiquitin unlocks autoinhibition of the catalytic cysteine. Our results show that PINK1-dependent phosphorylation of both parkin and ubiquitin is sufficient for full activation of parkin E3 activity. These findings demonstrate that phosphorylated ubiquitin is a parkin activator.

Dysfunction of PINK1 or parkin seems to cause an accumulation of low-quality depolarized mitochondria and production of excessive reactive oxygen species (ROS), which trigger familial Parkinsonism<sup>3–5</sup>. One of the most poorly understood events in PINK1- and parkin-mediated mitochondrial quality control is how the E3 activity of parkin is accelerated by damaged mitochondria. The simplest model is that PINK1-dependent phosphorylation of parkin accelerates its enzymatic activity. Indeed, parkin is phosphorylated at Ser 65 in a PINK1-dependent manner<sup>13–15</sup>. Furthermore, a phosphorylation-deficient mutation (Ser65Ala) of parkin hindered formation of the ubiquitin-ester intermediate necessary for parkin activation<sup>15</sup>. However, the requirement of PINK1 for parkin activation was not bypassed by the phosphomimetic parkin (Ser65Glu) (the mutation did not make PINK1 dispensable). We thus speculated that parkin is not the sole PINK1 substrate and that phosphorylation of other PINK1 substrate(s) is imperative for parkin activation.

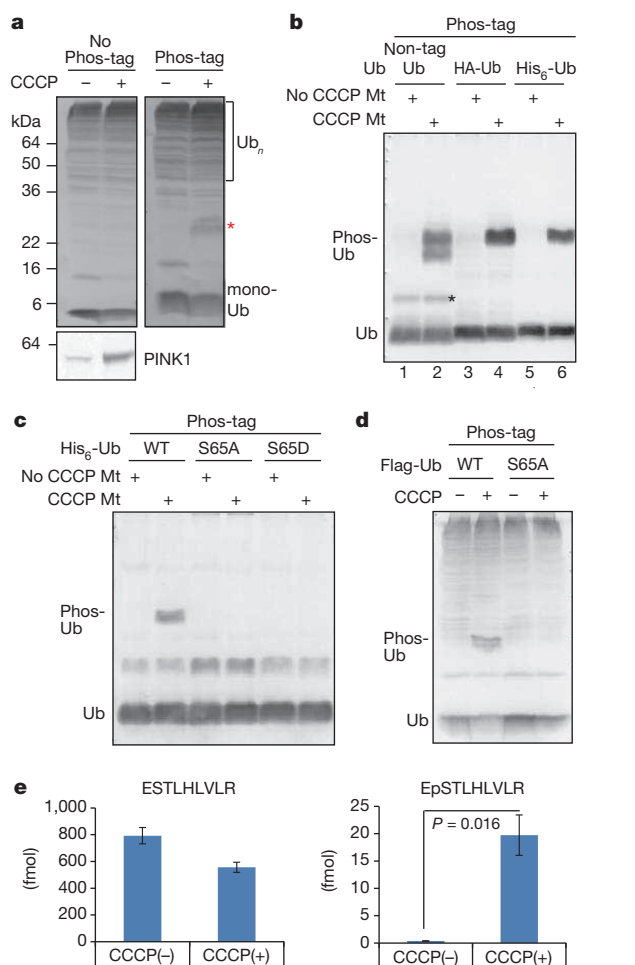
The structure of the amino-terminal parkin UBL (ubiquitin-like) domain resembles that of ubiquitin, and Ser 65 is conserved in both proteins (Extended Data Fig. 1a). Moreover, physical interactions between parkin and ubiquitin have been suggested<sup>16,17</sup>. We therefore examined whether ubiquitin is phosphorylated by PINK1. To detect a phosphorylated ubiquitin, we performed phosphate-affinity (Phos-tag) PAGE in which the phosphorylated ubiquitin can be easily distinguished from the non-phosphorylated form as a slower migrating band<sup>18</sup>. When PINK1-expressing HeLa-cell lysates treated with the protonophore carbonyl cyanide 3-chlorophenylhydrazone (CCCP) were subjected to Phos-tag PAGE, one retarded-mobility form of ubiquitin was specifically observed in the CCCP-treated cell lysate (Fig. 1a). When recombinant ubiquitin and ubiquitin with small N-terminal epitope tags were incubated with isolated mitochondria from CCCP-treated or untreated cells, and then subjected to Phos-tag PAGE (Fig. 1b and Extended Data Fig. 1b), ubiquitin phosphorylation was specifically observed in the reaction containing CCCP-pretreated mitochondria (Fig. 1b, lanes 2, 4, 6). In contrast, the small ubiquitin-related modifier-1 (SUMO-1) protein was not phosphorylated by depolarized mitochondria (Extended Data Fig. 1c), suggesting that phosphorylation of ubiquitin was specific.

To determine the phosphorylation site, the phosphorylated ubiquitin was trypsinized and subjected to liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis. Two phosphorylated ubiquitin peptides, 55–72 (TLSDYNIQKpSTLHLVLR) and 64–72 (EpSTLHLVLR) were identified with high confidence, suggesting Ser 65 phosphorylation (Extended Data Fig. 1d). Indeed, mutation of Ser 65 prevents ubiquitin phosphorylation by damaged mitochondria *in vitro* (Fig. 1c and Extended Data Fig. 1e) and following protonophore treatment in cells (Fig. 1d and Extended Data Fig. 1f). Next, the absolute levels of phosphorylated ubiquitin of PINK1-expressing HeLa cell lysates (Extended Data Fig. 1g, h) were determined using absolute quantification (AQUA) peptides as standards (Extended Data Table 1). The non-phosphorylated ESTLHLVLR peptide derived from endogenous ubiquitin was detected both in the presence and absence of CCCP, whereas the Ser 65 phosphopeptide (EpSTLHLVLR) was only detected in CCCP-pretreated cells (Fig. 1e and Extended Data Fig. 1i). Absolute quantification revealed that approximately 3% of total endogenous ubiquitin is phosphorylated in PINK1-expressing cells following a decrease in mitochondrial membrane potential ( $\Delta\Psi_m$ ).

We next investigated whether PINK1 is required for ubiquitin phosphorylation. When mitochondria derived from PINK1-knockout (*PINK1*<sup>−/−</sup>) mouse embryonic fibroblasts (MEFs)<sup>19</sup> were used (Fig. 2a and Extended Data Fig. 2a), ubiquitin phosphorylation was completely impeded even following CCCP pretreatment (Fig. 2a, lane 2). Wild-type PINK1 activated by CCCP treatment<sup>20,21</sup> complemented the phosphorylation of ubiquitin (Fig. 2a, lane 4), whereas PINK1 mutants with decreased kinase

<sup>1</sup>Laboratory of Protein Metabolism, Tokyo Metropolitan Institute of Medical Science, Setagaya-ku, Tokyo 156-8506, Japan. <sup>2</sup>Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-8561, Japan. <sup>3</sup>Division of Cell Signaling, Fujii Memorial Institute of Medical Sciences, The University of Tokushima, Tokushima 770-8503, Japan. <sup>4</sup>Research Center for Materials Science, Nagoya University, Nagoya, Aichi 464-8602, Japan. <sup>5</sup>Graduate School of Agriculture, Shizuoka University, 836 Ohya, Shizuoka 422-8529, Japan. <sup>6</sup>Molecular Profiling Research Center for Drug Discovery, National Institute of Advanced Industrial Science and Technology, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan. <sup>7</sup>JST-CREST/Department of Chemistry, Graduate School of Science, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan. <sup>8</sup>JST-CREST/Faculty of Life Sciences, Kyoto Sangyo University, Kamigamo-motoyama, Kita-ku, Kyoto 603-8555, Japan. <sup>9</sup>McGill Parkinson Program, Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, Montreal, Québec H3A 2B4, Canada. <sup>10</sup>Department of Pharmacology & Therapeutics, McGill University, Montreal, Québec H3G 1Y6, Canada. <sup>11</sup>Protein Metabolism Project, Tokyo Metropolitan Institute of Medical Science, Setagaya-ku, Tokyo 156-8506, Japan.

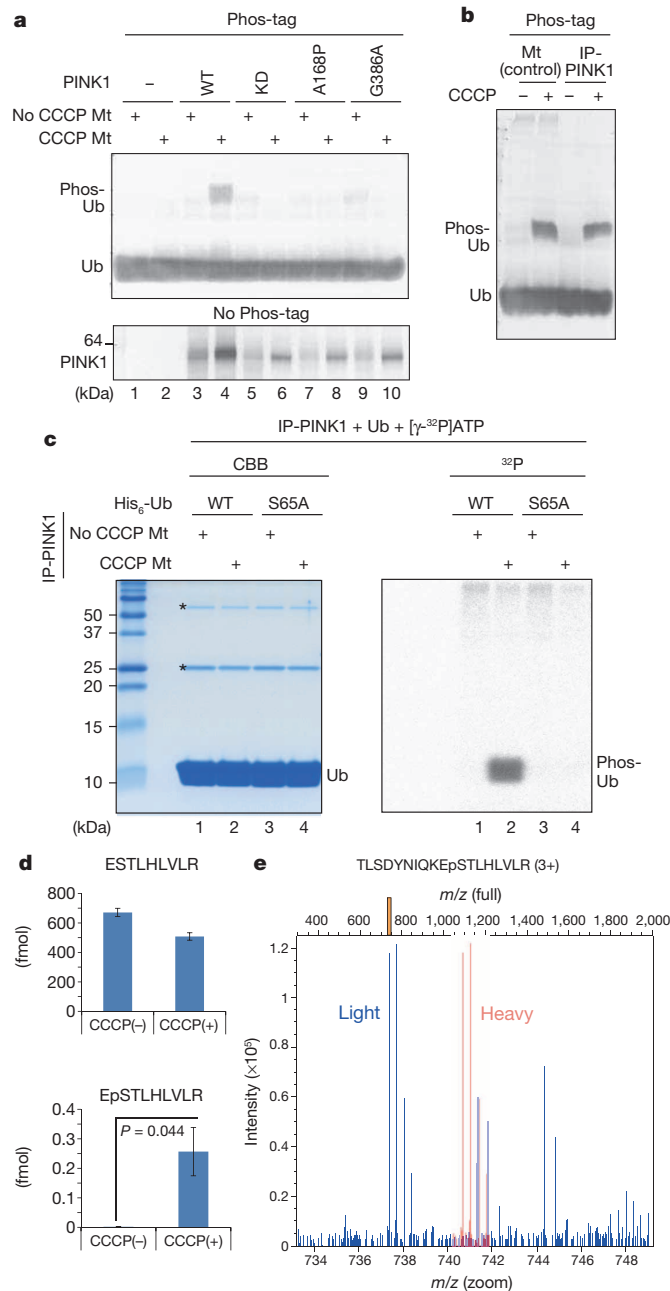




**Figure 1 | Ubiquitin Ser 65 is phosphorylated when  $\Delta\Psi_m$  is decreased.**

**a**, HeLa cells stably expressing PINK1 were treated with protonophore (CCCP) followed by Phos-tag PAGE and immunoblot analysis with an anti-ubiquitin (Ub) antibody. Red asterisk, phosphorylated ubiquitin. Ub<sub>n</sub>, ubiquitinated proteins. **b**, Recombinant ubiquitin was phosphorylated by depolarized mitochondria prepared from PINK1-expressing HeLa cells. Black asterisk, cross-reacting band; HA, haemagglutinin; His<sub>6</sub>, tandem repeat ( $\times 6$ ) of His residues; Phos-Ub, phosphorylated ubiquitin; Non-tag Ub, non-tagged ubiquitin. **c**, **d**, Mutation of Ser 65 prevents ubiquitin phosphorylation by depolarized mitochondria *in vitro* (**c**) and following protonophore treatment in cells (**d**). **e**, Absolute quantification of ESTLHLVLR and EpSTLHLVLR peptides in low-molecular-weight fraction of PINK1-expressing cells. Error bars represent mean  $\pm$  s.e.m. from three experiments. Statistical significance was calculated using a one-tailed paired *t*-test (significant if  $P < 0.05$ ). WT, wild type.

activity (kinase-dead, Ala168Pro, and Gly386Ala)<sup>20</sup> did not support phosphorylation of ubiquitin (Fig. 2a, lanes 6, 8, 10). To demonstrate PINK1-catalysed ubiquitin phosphorylation more convincingly, immunoprecipitated PINK1 (Extended Data Fig. 2b) was tested *in vitro*. Ubiquitin phosphorylation following incubation with the immunoprecipitated PINK1 was equivalent to that of CCCP-pretreated mitochondria (Fig. 2b and Extended Data Fig. 2c). Moreover, when recombinant ubiquitin was incubated with immunoprecipitated PINK1 and  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ , incorporation of  $^{32}\text{P}$  was detected in the recombinant wild-type ubiquitin (Fig. 2c, lane 2) but not the Ser65Ala ubiquitin mutant (Fig. 2c, lane 4). Taken together, these results demonstrate that PINK1 can phosphorylate ubiquitin at Ser 65. Because the data presented above were derived from PINK1-overproducing cells, we next quantified ubiquitin phosphorylation in intact HeLa cells (Fig. 2d and Extended Data Fig. 3) and found that 0.05% of total endogenous ubiquitin is phosphorylated. Although the absolute level was very low, given the pleiotropic functions of ubiquitin, such a small fraction might be a significant pool dedicated to the PINK1-mediated



**Figure 2 | PINK1 phosphorylates ubiquitin.** **a**, Active PINK1 is essential for ubiquitin phosphorylation *in vitro*. Mitochondria isolated from PINK1-knockout (*PINK1*<sup>-/-</sup>) mouse embryonic fibroblasts (MEFs) transfected with wild-type and mutant PINK1 were used. **b**, Immunoprecipitated PINK1 (IP-PINK1) can phosphorylate ubiquitin *in vitro* to the same level as depolarized mitochondria. **c**, Ubiquitin was labelled with  $^{32}\text{P}$  by IP-PINK1 prepared from CCCP-treated mitochondria, whereas the ubiquitin(Ser65Ala) mutant was not. Asterisks, heavy and light chains of immunoglobulin. **d**, Absolute quantification of endogenous phosphorylated ubiquitin in intact HeLa cells. Statistical analysis was performed as in Fig. 1e. **e**, Precursor ( $M + 3H$ )<sup>3+</sup> spectrum of the peptide harbouring phosphorylated Ser 65. The mass spectrometry peaks are shown in blue, and the orange bar indicates the position of the precursor. The theoretical position of the heavy precursor peptide is shown in transparent red.

activation of parkin. We tested further the dependency on endogenous PINK1. HeLa cells grown in light or heavy SILAC (stable isotope labeling using amino acids in cell culture) media were treated with non-target or PINK1 siRNA, respectively (Extended Data Fig. 2d), and then subjected to LC-MS/MS analysis. Phosphorylated Ser 65 peptide was present in control siRNA cells (Fig. 2e, peak identified as 'light') but absent

in *PINK1* siRNA (siPINK1) cells (peak identified as 'heavy'), indicating endogenous PINK1 phosphorylated ubiquitin at Ser 65 in intact cells.

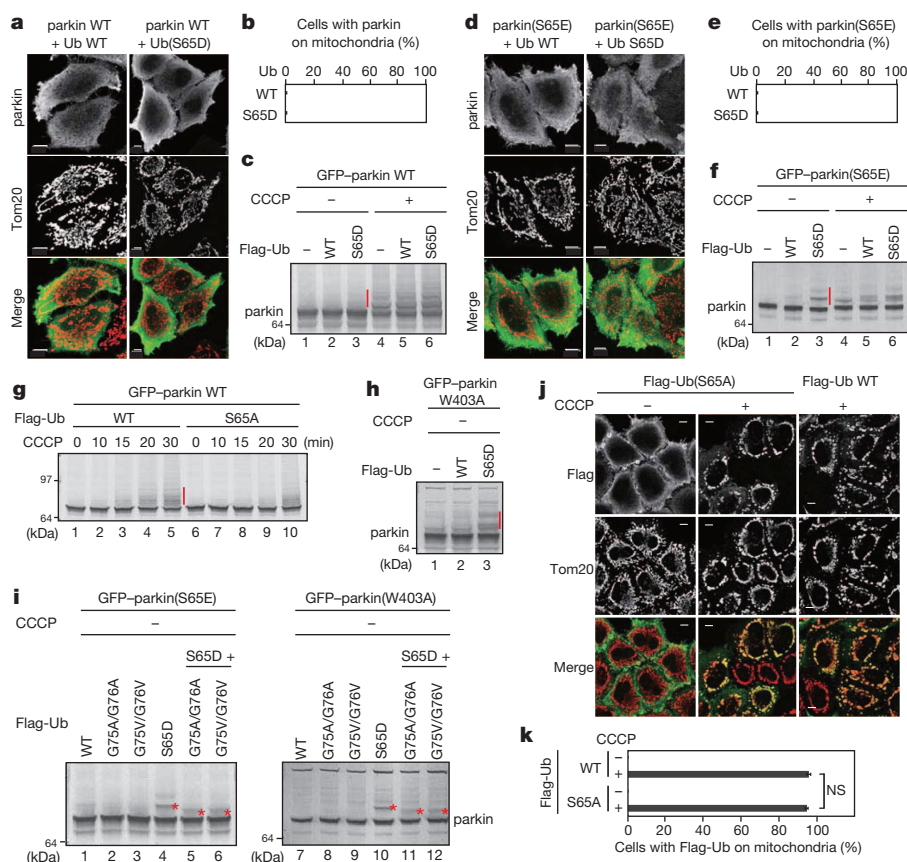
We next investigated the role of phosphorylated ubiquitin. If ubiquitin is the genuine PINK1 substrate, a phosphomimetic ubiquitin mutant would be expected to bypass the PINK1-dependency of parkin activation<sup>8</sup> or parkin recruitment onto depolarized mitochondria<sup>8–12</sup>. To monitor the E3 activity of parkin, autoubiquitination of green fluorescent protein (GFP)-tagged parkin (GFP-parkin) was used as an index<sup>8</sup>. In our first trial, expression of phosphomimetic ubiquitin (Ser65Asp) mutants led to neither mitochondrial localization (Fig. 3a, b) nor autoubiquitination of GFP-parkin (Fig. 3c, lane 3). However, the CCCP treatment triggered both mitochondrial localization (not shown) and autoubiquitination (Fig. 3c, lanes 4–6) of GFP-parkin. Because PINK1 also phosphorylates parkin<sup>13–15</sup>, we used a phosphomimetic parkin mutant, parkin(Ser65Glu)<sup>15</sup>, in case both phosphorylation events are necessary. Cytoplasmic localization of GFP-parkin(Ser65Glu) remained unchanged following co-expression with the ubiquitin(Ser65Asp) (Fig. 3d, e), whereas CCCP treatment stimulated translocation of GFP-parkin(Ser65Glu) to mitochondria (not shown). In sharp contrast, phosphomimetic ubiquitin (Ser65Asp) localized in cytoplasm (Extended Data Fig. 4a) promoted autoubiquitination of GFP-parkin(Ser65Glu) even in the absence of CCCP treatment (Fig. 3f, lane 3). Moreover, overproduction of phosphorylation-deficient ubiquitin (Ser65Ala) delayed CCCP-triggered parkin activation as compared with phosphorylatable wild-type ubiquitin, suggesting that the unphosphorylatable ubiquitin mutant competed with endogenous ubiquitin and thus exerted a dominant-negative effect (Fig. 3g).

Recently, it was noted that the parkin Trp403Ala mutation weakened autoinhibition and accelerated E3 activity<sup>22</sup>. When GFP-parkin (Trp403Ala) was co-expressed with ubiquitin(Ser65Asp), its autoubiquitination was also observed even without CCCP treatment (Fig. 3h, lane 3). These results suggest phosphorylated ubiquitin functions as a parkin activator; otherwise parkin might preferentially use phosphorylated ubiquitin for conjugation. To investigate this, we mutated the carboxy-terminal

diglycine motif of ubiquitin (Ser65Asp), which is crucial for formation of the thioester intermediate during the ubiquitination cascade. As shown in Fig. 3i, phosphomimetic ubiquitin with the diglycine mutations (Gly75Val/Gly76Val or Gly75Ala/Gly76Ala) still stimulated the E3 activity of GFP-parkin(Ser65Glu) and GFP-parkin(Trp403Ala) (Fig. 3i, lanes 5, 6, 11, 12; endogenous ubiquitin was conjugated). Moreover, parkin catalyses conjugation of phosphorylation-deficient ubiquitin (Ser65Ala) to depolarized mitochondria following CCCP treatment (Fig. 3j, k), indicating that parkin does not exclusively use phosphorylated ubiquitin for ubiquitination. Nevertheless, phosphomimetic ubiquitin could be recruited to mitochondria and conjugated to mitofusin-2 (Extended Data Fig. 4b, c) in a CCCP-dependent manner. Thus, parkin conjugates both unphosphorylated and phosphorylated ubiquitin to substrates.

We confirmed further the aforementioned results in a cell-free assay using GFP-parkin and purified mitochondria<sup>15,23</sup>. Autoubiquitination of GFP-parkin in intact cell extracts was specifically observed when incubated with CCCP-pretreated mitochondria<sup>15,23</sup>. In contrast, autoubiquitination of GFP-parkin mutants harbouring the Ser65Glu or Trp403Ala mutations was clearly observed when incubated with recombinant phosphomimetic ubiquitin (Ser65Asp) in the absence of depolarized mitochondria (Fig. 4a, lanes 12, 18). Wild-type ubiquitin and the phosphorylation-deficient Ser65Ala mutant did not support autoubiquitination of parkin(Ser65Glu) or parkin(Trp403Ala) (lanes 7–10, 13–16), and ubiquitin(Ser65Asp) did not accelerate autoubiquitination of wild-type GFP-parkin (lane 6). To rule out the possible effect of PINK1 contamination, we repeated the experiments using GFP-parkin prepared from *PINK1*<sup>−/−</sup> MEFs. Autoubiquitination of both Ser65Glu and Trp403Ala mutants of GFP-parkin were again observed (Fig. 4b), further supporting the idea that phosphomimetic ubiquitin makes PINK1 dispensable for phosphomimetic parkin activation.

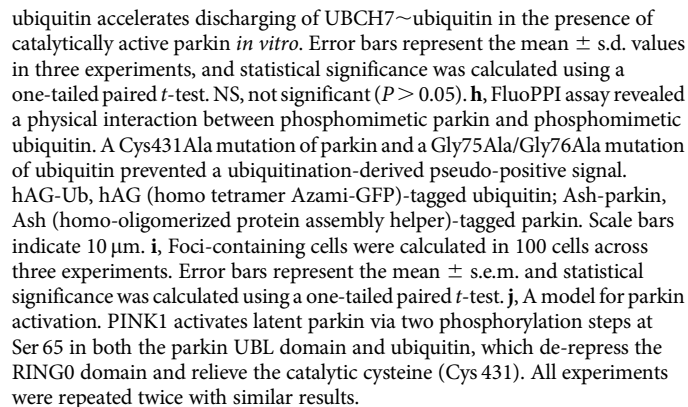
We next examined directly whether phosphorylated ubiquitin (not phosphomimetic ubiquitin) activated parkin. Ubiquitin was phosphorylated by incubation with depolarized mitochondria, then subjected to



**Figure 3 | Phosphomimetic ubiquitin activates phosphomimetic parkin.**

**a**, Cytoplasmic localization of WT parkin co-expressed with WT or phosphomimetic ubiquitin (Ser65Asp). Scale bars indicate 10  $\mu$ m. **b**, The rate of mitochondrial parkin localization was 0% when calculated in 100 cells across three experiments. **c**, Autoubiquitination of GFP-parkin was not observed by co-expression of phosphomimetic ubiquitin (red vertical bars). **d**, **e**, Cytoplasmic localization of the phosphomimetic parkin(Ser65Glu) mutant co-expressed with WT or phosphomimetic ubiquitin. Immunocytochemistry (**d**) and statistical analysis (**e**) were performed as in **a** and **b**. **f**, Autoubiquitination of phosphomimetic GFP-parkin(Ser65Glu) was observed following co-expression of phosphomimetic ubiquitin (Ser65Asp) even in the absence of CCCP treatment (lane 3). **g**, Co-expression of phosphorylation-deficient ubiquitin (Ser65Ala) mutant delayed CCCP-dependent activation of WT parkin. **h**, Phosphomimetic ubiquitin (Ser65Asp) induced autoubiquitination of GFP-parkin(Trp403Ala). **i**, The C-terminal diglycine motif (GG) is dispensable for parkin activation. Red asterisks, autoubiquitination of the GFP-parkin(Ser65Glu) and GFP-parkin(Trp403Ala) mutants. **j**, Mitochondrial localization of phosphorylation-deficient ubiquitin (Ser65Ala) mediated by parkin, following a decrease in  $\Delta\Psi$ m. **k**, The rate of cells with Flag-Ub on mitochondria was calculated as a percentage using 100 cells across three experiments. Error bars represent the mean  $\pm$  s.e.m. NS, not significant ( $P > 0.05$ ) in a one-tailed Student's *t*-test.

Finally, we sought to define the molecular mechanism by which phosphorylated ubiquitin activates parkin. We confirmed that phosphorylated ubiquitin was charged on UBC7 as readily as non-phosphorylated ubiquitin (Fig. 4f and Extended Data Fig. 7a), indicating a similar efficiency





of conjugation by E2. We then examined whether phosphomimetic and phosphorylated ubiquitin accelerates discharge of UBCH7~unmodified ubiquitin in the presence of parkin *in vitro* (note that the reaction does not contain ATP, thus free phosphorylated ubiquitin cannot enter into the conjugation pathway). Both phosphomimetic and phosphorylated ubiquitin accelerates discharge of ubiquitin from UBCH7~ubiquitin depending on the presence of catalytic cysteine (Cys 431) of parkin, revealing that the activation mechanism is strictly allosteric (Fig. 4g and Extended Data Fig. 7b). We next examined whether phosphorylated ubiquitin interacts physically with parkin. Using the FluorPPI (fluorescent-based technology detecting protein–protein interactions) technique, in which protein–protein interactions are detectable as foci formation in living cells, we demonstrated a clear interaction between parkin and ubiquitin in cells. Importantly, parkin with an intact UBL domain failed to interact with phosphomimetic ubiquitin, whereas parkin with a phosphomimetic (Ser65Glu) UBL mutation was able to interact (Figs. 4h, i). This result suggests that the interaction between parkin and phosphorylated ubiquitin competes with the intact parkin UBL domain, whereas phosphorylation of the UBL domain releases this inhibitory effect. Recently, the structure of parkin<sup>22,26–28</sup> revealed that the RINGO domain occludes the catalytic cysteine (Cys 431). To gain further insight into the mechanism, we used computational modelling to generate two structural models that support binding of phosphorylated ubiquitin or phosphorylated UBL domain facilitates accessibility to Cys 431 (Extended Data Fig. 8). Because phosphorylation at both sites is required for parkin activation (Fig. 3), either of these conformations might be an intermediate to the other. Based on our combined results and the autoinhibitory effect of the UBL domain<sup>16</sup>, we speculate that phosphorylation of both ubiquitin and the UBL domain unlock repression of the catalytic cysteine by the RINGO domain, thereby converting parkin to its fully-active form (Fig. 4j).

In the present study, we demonstrate—to our knowledge, for the first time—that parkin activation in response to mitochondrial damage proceeds following ubiquitin phosphorylation by PINK1. The E3 activity of parkin is usually repressed via autoinhibition<sup>8,16,22,26,27</sup> but is specifically liberated by PINK1 in a unique two-step phosphorylation-dependent manner: one step is the phosphorylation of the UBL domain of parkin<sup>13–15</sup>, and the other step is, unexpectedly, phosphorylation of ubiquitin (determined in this work). As mitochondrial dysfunction has long been associated with sporadic Parkinson's disease<sup>29</sup>, a complete understanding of the molecular mechanism underlying parkin activation is expected to clarify the pathogenesis of not only hereditary but also sporadic forms of Parkinson's disease. Moreover, although thousands of papers studying ubiquitin have been published, this work is the first, to our knowledge, to reveal ubiquitin phosphorylation and its significance.

## METHODS SUMMARY

Immunoprecipitation of PINK1 (ref. 21), Phos-tag PAGE<sup>15,20</sup>, *in vitro* ubiquitination of GFP–parkin on depolarized mitochondria<sup>15,23</sup>, charging and discharging assay<sup>22</sup>, and mass spectrometric analyses<sup>15,20,30</sup> were performed as reported previously. Detailed procedures are available in the Methods (online).

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 January; accepted 23 April 2014.

Published online 30 April 2014.

1. Kitada, T. *et al.* Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605–608 (1998).
2. Valente, E. M. *et al.* Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science* **304**, 1158–1160 (2004).
3. Narendra, D., Walker, J. E. & Youle, R. Mitochondrial quality control mediated by PINK1 and Parkin: links to Parkinsonism. *Cold Spring Harb. Perspect. Biol.* **4**, a011338 (2012).
4. Corti, O., Lesage, S. & Brice, A. What genetics tells us about the causes and mechanisms of Parkinson's disease. *Physiol. Rev.* **91**, 1161–1218 (2011).

5. Exner, N., Lutz, A. K., Haass, C. & Winklhofer, K. F. Mitochondrial dysfunction in Parkinson's disease: molecular mechanisms and pathophysiological consequences. *EMBO J.* **31**, 3038–3062 (2012).
6. Clark, I. E. *et al.* *Drosophila pink1* is required for mitochondrial function and interacts genetically with parkin. *Nature* **441**, 1162–1166 (2006).
7. Park, J. *et al.* Mitochondrial dysfunction in *Drosophila PINK1* mutants is complemented by parkin. *Nature* **441**, 1157–1161 (2006).
8. Matsuda, N. *et al.* PINK1 stabilized by mitochondrial depolarization recruits Parkin to damaged mitochondria and activates latent Parkin for mitophagy. *J. Cell Biol.* **189**, 211–221 (2010).
9. Narendra, D. P. *et al.* PINK1 is selectively stabilized on impaired mitochondria to activate Parkin. *PLoS Biol.* **8**, e1000298 (2010).
10. Geisler, S. *et al.* PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *Nature Cell Biol.* **12**, 119–131 (2010).
11. Vives-Bauza, C. *et al.* PINK1-dependent recruitment of Parkin to mitochondria in mitophagy. *Proc. Natl Acad. Sci. USA* **107**, 378–383 (2010).
12. Ziviani, E., Tao, R. N. & Whitworth, A. J. *Drosophila* parkin requires PINK1 for mitochondrial translocation and ubiquitinates mitofusin. *Proc. Natl Acad. Sci. USA* **107**, 5018–5023 (2010).
13. Kondapalli, C. *et al.* PINK1 is activated by mitochondrial membrane potential depolarization and stimulates Parkin E3 ligase activity by phosphorylating Serine 65. *Open Biol.* **2**, 120080 (2012).
14. Shiba-Fukushima, K. *et al.* PINK1-mediated phosphorylation of the Parkin ubiquitin-like domain primes mitochondrial translocation of Parkin and regulates mitophagy. *Sci. Rep.* **2**, 1002 (2012).
15. Iguchi, M. *et al.* Parkin-catalyzed ubiquitin-ester transfer is triggered by PINK1-dependent phosphorylation. *J. Biol. Chem.* **288**, 22019–22032 (2013).
16. Chaugule, V. K. *et al.* Autoregulation of Parkin activity through its ubiquitin-like domain. *EMBO J.* **30**, 2853–2867 (2011).
17. Zheng, X. & Hunter, T. Parkin mitochondrial translocation is achieved through a novel catalytic activity coupled mechanism. *Cell Res.* **23**, 886–897 (2013).
18. Kinoshita, E., Kinoshita-Kikuta, E., Takiyama, K. & Koike, T. Phosphate-binding tag, a new tool to visualize phosphorylated proteins. *Mol. Cell. Proteom.* **5**, 749–757 (2006).
19. Gautier, C. A., Kitada, T. & Shen, J. Loss of PINK1 causes mitochondrial functional defects and increased sensitivity to oxidative stress. *Proc. Natl Acad. Sci. USA* **105**, 11364–11369 (2008).
20. Okatsu, K. *et al.* PINK1 autophosphorylation upon membrane potential dissipation is essential for Parkin recruitment to damaged mitochondria. *Nature Commun.* **3**, 1016 (2012).
21. Okatsu, K. *et al.* A dimeric PINK1-containing complex on depolarized mitochondria stimulates Parkin recruitment. *J. Biol. Chem.* **288**, 36372–36384 (2013).
22. Trempe, J. F. *et al.* Structure of parkin reveals mechanisms for ubiquitin ligase activation. *Science* **340**, 1451–1455 (2013).
23. Lazarou, M. *et al.* PINK1 drives Parkin self-association and HECT-like E3 activity upstream of mitochondrial binding. *J. Cell Biol.* **200**, 163–172 (2013).
24. Ciechanover, A., Hod, Y. & Herschko, A. A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. 1978. *Biochem. Biophys. Res. Commun.* **81**, 1100–1105 (1978).
25. Spence, J., Sadis, S., Haas, A. L. & Finley, D. A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol. Cell. Biol.* **15**, 1265–1273 (1995).
26. Riley, B. E. *et al.* Structure and function of Parkin E3 ubiquitin ligase reveals aspects of RING and HECT ligases. *Nature Commun.* **4**, 1982 (2013).
27. Wauer, T. & Komander, D. Structure of the human Parkin ligase domain in an autoinhibited state. *EMBO J.* **32**, 2099–2112 (2013).
28. Spratt, D. E. *et al.* A molecular explanation for the recessive nature of parkin-linked Parkinson's disease. *Nature Commun.* **4**, 1983 (2013).
29. Schapira, A. H. Complex I: inhibitors, inhibition and neurodegeneration. *Exp. Neurol.* **224**, 331–335 (2010).
30. Tsuchiya, H., Tanaka, K. & Saeki, Y. The parallel reaction monitoring method contributes to a highly sensitive polyubiquitin chain quantification. *Biochem. Biophys. Res. Commun.* **436**, 223–229 (2013).

**Acknowledgements** This work was supported by JSPS KAKENHI grant numbers 23687018 (to N.M.), 23-6061 (to K.O.), 21000012 (to K.T.), 24657072 and 22227003 (to Y.T. and T.E.); MEXT KAKENHI grant numbers 24111557 and 25112522 (to N.M.), 24112008 (to Y.S.); the Tomizawa Jun-ichi and Keiko Fund (to N.M.); JST CREST (to T.E.); Platform for Drug Discovery, Informatics, and Structural Life Science from MEXT (to T.H.); startup funds from McGill University and the GRASP from the FRSQ (to J.-F.T.); an Operating Grant from the CIHR and a Chercheur National Award from the FRSQ (to E.A.F.); and the Takeda Science Foundation (to H.K., K.T. and N.M.). We thank N. Croteau and S. Rasool for technical assistance, T. Ueno for anti-FoF1 antibody, and D. Finley for yeast strain SUB328.

**Author Contributions** F.K., K.O., H.K., E.G., J.-F.T. and M.K. carried out biochemical and immunoblot experiments. T.H. performed computational modelling. Y.T., Y.K., Y.S. and T.E. performed yeast experiments. F.K., H.T., H.Y., J.-F.T., E.A.F. and Y.S. performed mass spectrometric analysis. K.T. and N.M. designed the research and analysed the data. N.M. wrote the manuscript with help and supervision from K.T. All authors contributed to data analysis and preparation of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Noriyuki Matsuda ([matsuda-nr@igakuken.or.jp](mailto:matsuda-nr@igakuken.or.jp)) and Keiji Tanaka ([tanaka-kj@igakuken.or.jp](mailto:tanaka-kj@igakuken.or.jp)).

## METHODS

**Cells, plasmids and transfections.** HeLa cells and MEFs were cultured at 32–37 °C with 5% CO<sub>2</sub> in Dulbecco's modified Eagle's medium (DMEM, Sigma) containing 1× nonessential amino acids (Gibco), 1× sodium pyruvate (Gibco), and 10% fetal bovine serum (Gibco). *PINK1*<sup>−/−</sup> MEFs stably expressing wild-type or mutant *PINK1* were established as reported previously<sup>8</sup>. To generate HeLa cells stably expressing *PINK1*-3×Flag, HeLa cells transiently expressing mCAT1 (murine cationic amino acid transporter-1) were infected with recombinant retroviruses harbouring *PINK1*-3×Flag. Recombinant retrovirus was made using PLAT-E cells as reported previously<sup>15</sup>. Plasmids for expressing wild-type or various mutants of parkin and ubiquitin were as described previously<sup>8,15</sup> or were newly constructed by conventional methods. Plasmid transfections were performed using the transfection reagent FuGene6 (Roche) for HeLa cells and polyethylenimine (Polyscience) for *PINK1*<sup>−/−</sup> MEFs. To depolarize the mitochondria, cells were treated with 15–30 μM CCCP (Wako) for 2–3 h, except in Fig. 3g, in which cells were treated with CCCP for 10–30 min. siRNA targeting *PINK1* (J-004030-0008-0002, GGACGCGUUC CUCGUUUAU) or non-targeting control (D-001810-01-05, mixture of UGGUUU ACAUUGCAGACUAA, UGGUUUACAUGUUGUGUGA, UGGUUUACAUGU UAUCUGA and UGGUUUACAUGUUUCCUA) were purchased from Dharmacon (Thermo Fisher Scientific).

**Phos-tag assay.** To detect phosphorylated proteins using SDS-PAGE, 12.5–15% polyacrylamide gels containing 50 μM Phos-tag acrylamide (Wako chemicals) and 100 μM MnCl<sub>2</sub> were used (refs 15, 20). After electrophoresis, Phos-tag acrylamide gels were washed with gentle shaking in transfer buffer containing 0.01% SDS and 1 mM EDTA for 10 min and then incubated in transfer buffer containing 0.01% SDS without EDTA for 10 min according to the manufacturer's protocol. Proteins were transferred to polyvinylidene difluoride (PVDF) membranes and studied by conventional immunoblot analysis as outlined below.

**Preparation of recombinant ubiquitin proteins.** For purification of recombinant His<sub>6</sub>-ubiquitin, *E. coli* Rosetta2 (DE3) (Novagen) bacterial cells transformed with plasmids encoding wild-type or mutant His<sub>6</sub>-ubiquitin were pre-cultured overnight at 37 °C in 20 ml of lysogeny broth (LB) medium supplemented with 100 μg ml<sup>−1</sup> ampicillin and 24 μg ml<sup>−1</sup> chloramphenicol, and then transferred to 200 ml of fresh medium. After incubation for 2 h at 37 °C, isopropyl-β-D-thiogalactoside (IPTG) was added at a final concentration of 1 mM, and *E. coli* were further cultured for 6 h at 37 °C. The cells were collected, suspended in 40 ml of 20 mM Tris-HCl, pH 7.5, and lysed by sonication. After centrifugation at 8,000 r.p.m. for 10 min, the supernatant was recovered and purified by conventional methods, dialysed against buffer A (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, and 10% glycerol) and stored at −80 °C. Non-tagged ubiquitin, HA-ubiquitin, His<sub>6</sub>-ubiquitin harbouring C-terminal Gly75Ala/Gly76Ala mutation, and SUMO-1 were purchased from Boston Biochem.

**Immunoblot analysis and immunofluorescence.** To observe autoubiquitination of GFP-parkin (an indicator for the re-establishment of the latent E3 activity of parkin) in immunoblot analyses, lysates of HeLa cells or MEFs were collected in TNE-N<sup>+</sup> buffer (150 mM NaCl, 20 mM Tris-HCl, pH 8.0, 1 mM EDTA and 1% NP-40) in the presence of 10 mM *N*-ethyl maleimide (NEM), which protects ubiquitinated proteins from deubiquitinase activity<sup>8</sup>. For phosphorylation immunoblot analyses, lysates from MEFs and HeLa cells were collected in the presence of PhosSTOP (Roche) to inhibit phosphatase activity. To detect various proteins via immunoblot analyses, the anti-ubiquitin antibody P4D1 (Santa Cruz, 1:1,000), Z0458 (DAKO, 1:500), anti-SUMO-1 (GMP-1) antibody 21C7 (ZYMED, 1:500), anti-parkin antibody PRK8 (Sigma, 1:2,000), CS2132 (MBL, 1:1,000), anti-Flag antibody 2H8 (Transgenic, 1:500), anti-Mfn2 antibody ab56889 (Abcam, 1:500), anti-PINK1 antibody BC100-494 (Novus, 1:1,000), anti-GFP antibody ab6556 (Abcam, 1:500), anti-VDAC1 antibody ab2 (Calbiochem, 1:1,000) and anti-FoF1 antibody (gift from T. Ueno, 1:1,000) were used.

To investigate the requirement of diglycine motif of phosphomimetic ubiquitin, we replaced the C-terminal diglycine motif with either Val or Ala (Gly75Val/Gly76Val and Gly75Ala/Gly76Ala) and coexpressed with GFP-parkin(Ser65Glu) or GFP-parkin(Trp403Ala). Note that autoubiquitination bands of GFP-parkin are shifted slightly upwards in the lane by conjugation of exogenous Flag-ubiquitin(Ser65Asp) when Flag-ubiquitin(Ser65Asp) and GFP-parkin were co-expressed. In contrast, autoubiquitination bands of GFP-parkin lie further down the lane when it is co-expressed with Flag-tagged phosphomimetic ubiquitin lacking a C-terminal diglycine motif; that is, only endogenous ubiquitin (not Flag-tagged phosphomimetic ubiquitin) can conjugate to GFP-parkin.

For immunofluorescence experiments, HeLa cells were fixed with 4% paraformaldehyde, permeabilized with 50 μg ml<sup>−1</sup> digitonin, and stained with a primary antibody (anti-GFP antibody ab6556 (Abcam, 1:500), anti-Flag antibody 2H8 (Transgenic, 1:500), anti-Tom20 antibody FL-145 (Santa Cruz Biotech., 1:3,000)) and a 1:2,000 dilution of the secondary antibody (Alexa Fluor 488- or 568-conjugated anti-mouse or -rabbit IgG antibody, (Invitrogen)). Cells were imaged using a laser-scanning microscope (LSM510; Carl Zeiss) and image brightness was adjusted in

Photoshop (Adobe). For statistical analysis, subcellular localization of parkin and ubiquitin were analysed in 100 cells across three experiments, and statistical significance was calculated using a one-tailed Student's *t*-test. When mitochondrial localization of phosphorylation-deficient ubiquitin(Ser65Ala) was examined by immunofluorescence, HeLa cells were treated with CCCP for 3 h to reach a plateau because overproduction of Ser65Ala delayed parkin activation.

**Cell-free reconstitution of GFP-parkin autoubiquitination.** HeLa cells expressing exogenous GFP-parkin were suspended in cell-free assay buffer (20 mM HEPES-KOH (pH 7.5), 220 mM sorbitol, 10 mM KAC, 70 mM sucrose) supplemented with a protease inhibitor cocktail minus EDTA (Roche). Cells were disrupted by passing 30 times through a 25-gauge needle, and cell homogenates were centrifuged at 800g for 10 min at 4 °C to obtain a post-nuclear supernatant. Cytosolic fractions were collected by further centrifugation at 20,400g for 10 min at 4 °C. For isolation of mitochondria, HeLa cells or MEFs expressing *PINK1*-3×Flag were treated with 15–30 μM CCCP for 3 h followed by homogenization in the aforementioned cell-free assay buffer. Post-nuclear supernatants were then obtained by centrifugation as described above, and mitochondria were pelleted by further centrifugation at 10,000g for 20 min at 4 °C. To perform the cell-free ubiquitination assay, HeLa cytosolic fractions containing exogenous GFP-parkin were incubated with isolated mitochondria supplemented with 5 mM MgCl<sub>2</sub>, 5 mM ATP, 2 mM DTT, and 1% glycerol at 30 °C for 2 h (refs 15, 23). To demonstrate that phosphorylated ubiquitin activates parkin, ubiquitin was phosphorylated beforehand using isolated mitochondria, subjected to ultracentrifugation at 20,400g for 10 min at 4 °C to deplete mitochondria, and treated at 90 °C for 10 min. HeLa cytosolic fractions containing exogenous GFP-parkin were incubated with non-phosphorylated or phosphorylated ubiquitin (final concentration is 50 μg ml<sup>−1</sup>) supplemented with 5 mM MgCl<sub>2</sub>, 5 mM ATP, 2 mM DTT, and 1% glycerol at 30 °C for 2 h.

**Immunoprecipitation of PINK1.** Immunoprecipitation of PINK1 was performed as reported previously (ref. 21), but with some modifications. For immunoprecipitation experiments, mitochondria from HeLa cells stably expressing *PINK1*-3×Flag were collected as described above, re-suspended in cell-free assay buffer (20 mM HEPES-KOH (pH 7.5), 220 mM sorbitol, 10 mM KAC, 70 mM sucrose), solubilized with 10 mg ml<sup>−1</sup> digitonin (Wako) for 15 min at 4 °C, and reacted with agarose (Protein G Sepharose 4 Fast Flow, GE Life Sciences) conjugated with anti-Flag antibody 2H8 for 1 h at 4 °C. The resulting immunoprecipitates were washed repeatedly with the same buffer and collected by centrifugation. Its purity was confirmed by immunoblot analysis with antibodies against other mitochondrial membrane proteins such as VDAC, mitofusin-2 and FoF1 ATPase.

**In vitro phosphorylation of ubiquitin.** To phosphorylate ubiquitin with depolarized mitochondria *in vitro*, CCCP-treated mitochondria were collected as described above and then incubated with recombinant His<sub>6</sub>-ubiquitin (final concentration of 40 ng μl<sup>−1</sup>) in cell-free assay buffer (20 mM HEPES-KOH (pH 7.5), 220 mM sorbitol, 10 mM KAC, 70 mM sucrose) supplemented with 5 mM MgCl<sub>2</sub>, 5 mM ATP, 2 mM DTT and 1% glycerol at 30 °C for 1 h. The supernatants were subsequently subjected to centrifugation at 16,000g for 10 min to deplete mitochondria. To phosphorylate ubiquitin with immunoprecipitated PINK1 *in vitro*, immunoprecipitated PINK1 was added to the aforementioned assay instead of depolarized mitochondria. For the *in vitro* kinase assay using [ $\gamma$ -<sup>32</sup>P]ATP, immunoprecipitated *PINK1*-3×Flag was incubated with HA-ubiquitin (10 μg) or His<sub>6</sub>-ubiquitin (wild-type or Ser65Ala; 30 μg) and 100 μM [ $\gamma$ -<sup>32</sup>P]ATP (5 μCi) in 30 μl kinase buffer (20 mM Tris-HCl, pH 7.5, 5 mM MgCl<sub>2</sub> and 1 mM DTT) for 30 min at 30 °C. The reaction was stopped by adding Laemmli's sample buffer and boiling. One-third of the sample was subjected to 17% SDS-PAGE and CBB staining. Phosphorylated proteins were then visualized by autoradiography.

**LC-MS/MS analysis of ubiquitin.** Mass spectrometric analyses were performed as reported previously (refs 15, 20, 30), but with some modification. To identify ubiquitin phosphorylation sites, recombinant ubiquitin was reacted with isolated mitochondria from CCCP-treated and untreated cells. Recombinant ubiquitin was then subjected to SDS-PAGE, stained with CBB, and in-gel trypsin digestion was carried out as reported previously<sup>30</sup>. Gels were extensively washed with Milli-Q water (Millipore), the bands of interest were excised, cut into 1-mm<sup>2</sup> pieces, and destained with 1 ml of 50 mM ammonium bicarbonate (AMBC) buffer containing 50% acetonitrile (ACN) with agitation for 1 h. A final 100% ACN wash was performed to ensure complete gel dehydration. Trypsinization solution (20 ng μl<sup>−1</sup>) was prepared by diluting modified sequencing grade trypsin (Promega) with 50 mM AMBC buffer, pH 8.0, containing 5% ACN. The trypsin solution was added to the gel pieces and incubated overnight at 37 °C. Digests were quenched and extracted by addition of 50 μl of 50% ACN containing 0.1% trifluoroacetic acid (TFA) mixture for 1 h by shaking. The digested peptides were recovered into fresh Protein LoBind tubes and an additional extraction was performed with 70% ACN containing 0.1% TFA mixture for 30 min. The extracted peptides were concentrated to 20 μl by speed-vac. The concentrated peptides were prepared in 0.1% TFA. The resultant peptides were analyzed on a nanoflow UHPLC instrument (Easy nLC 1000,

Q-Exactive MS and nano-electrospray ion source (Thermo Fisher Scientific)) with the raw data processed using Xcalibur (Thermo Fisher Scientific). MS spectra were analyzed using Protein Discoverer software version 1.3 (Thermo Fisher Scientific). The fragmentation spectra were searched against the UniProt database with the MASCOT search engine.

To test the dependency on endogenous PINK1, HeLa cells grown in light or heavy SILAC (stable isotope labeling using amino acids in cell culture) media were treated with non-target or *PINK1* siRNA, respectively. Precursor ( $M + 3H$ )<sup>3+</sup> spectrum of the phosphorylated Ser 65 peptide of each sample was detected by LC-MS/MS analysis.

To quantify the phosphorylation of endogenous ubiquitin following dissipation of  $\Delta\Psi_m$ , PINK1-expressing or intact HeLa cells were treated with CCCP. Whole-cell lysates were collected in TNE- $N^+$  lysis buffer, subjected to SDS-PAGE and stained with CBB. The low-molecular-weight fraction (3–14 k) of the gels (as shown in Fig. 1g) were excised and cut into 1-mm<sup>2</sup> pieces. In-gel digestion using trypsin and lysyl endopeptidase (Wako) was carried out as described above. The absolute level of phosphorylated ubiquitin in PINK1-expressing or intact HeLa cells was measured by parallel reaction monitoring (PRM), a MS/MS-based quantification method<sup>30</sup>. The extracted peptides were spiked with 10 fmol AQUA peptides as standards (ESTLHLVLR and EpSTLHLVLR, SIGMA) and analysed by Q Exactive in targeted MS/MS mode. Data were processed by PinPoint software version 1.3 (Thermo Fisher Scientific).

**Protein-protein interaction assay.** Physical interaction between phosphomimetic ubiquitin and phosphomimetic parkin was detected by FluorPPI (fluorescent-based technology detecting protein–protein interactions) system (Amalgam), in which the protein–protein interaction is detectable as the formation of fluorescent foci in living cells. The C-terminal diglycine motif in ubiquitin was mutated to prevent generation of a conjugation-derived pseudo-positive signal. This ubiquitin mutant was sub-cloned into pHAG-MCL plasmid to fuse with a hAG (homo tetramer Azami-Green fluorescent protein) tag. Cys 431 in parkin was likewise mutated to prevent an autoubiquitination-derived pseudo-positive signal. The mutant was sub-cloned into the pAsh-MCL plasmid to fuse with Ash (homo oligomerized protein assembly helper) tag. The interaction between the hAG-tagged protein and the Ash-tagged protein induces formation of green fluorescent foci in cells. Ash-tagged parkin (Cys431Ala) and hAG-tagged ubiquitin (Gly75Ala, Gly76Ala) were co-transfected in cells and intracellular foci were observed using confocal microscopy (LSM710, Zeiss). For statistical analysis, foci-containing cells were counted in 100 cells across three experiments, and statistical significance was calculated using one-tailed paired *t*-test.

**Charging and discharging assay of UBCH7.** The UBCH7 charging assay was performed as follows (ref. 22). We first incubated 40  $\mu$ M His-Ub (wild-type, Ser65Ala or Ser65Asp) with 2  $\mu$ M GST-TcPINK1 (ref. 13) and 2 mM Mg-ATP for 30 min, and then incubated the products (20  $\mu$ M His-Ub) with 4  $\mu$ M UBCH7, 0.1  $\mu$ M E1, 2 mM Mg-ATP for 20 min. The reaction was stopped with sample buffer, with or without 50 mM DTT. The samples were resolved on SDS-PAGE with or without 25  $\mu$ M Phos-tag-MnCl<sub>2</sub>, and gel was stained with Coomassie Blue. The results show clearly that UBCH7 can be charged with phosphorylated ubiquitin, as seen in the shift observed in UBCH7~pUb, UBCH7~phosphorylated ubiquitin or charging of Ser65Asp. UBCH7 discharging assay was performed as follows (ref. 22). First His-UBCH7 was charged with N-terminal Fluorescein-labelled ubiquitin (FluoUb), using E1 and Mg-ATP. His-UBCH7~FluoUb was purified by Ni-NTA agarose, and then His-UBCH7~FluoUb (1  $\mu$ M) was mixed with wild-type or mutant GST-parkin (0.5  $\mu$ M) in the presence of 10  $\mu$ M His-Ub wild-type, Ser65Asp or phosphorylated HisUb, generated *in vitro* using TcPINK1. Mixtures were incubated at 37 °C for 10 min, resolved on SDS-PAGE, and imaged using the Typhoon fluorescent scanner (GE). Fluorescence strength of His-UBCH7~FluoUb in three experiments were measured for accurate quantification, and were subjected to statistical analysis.

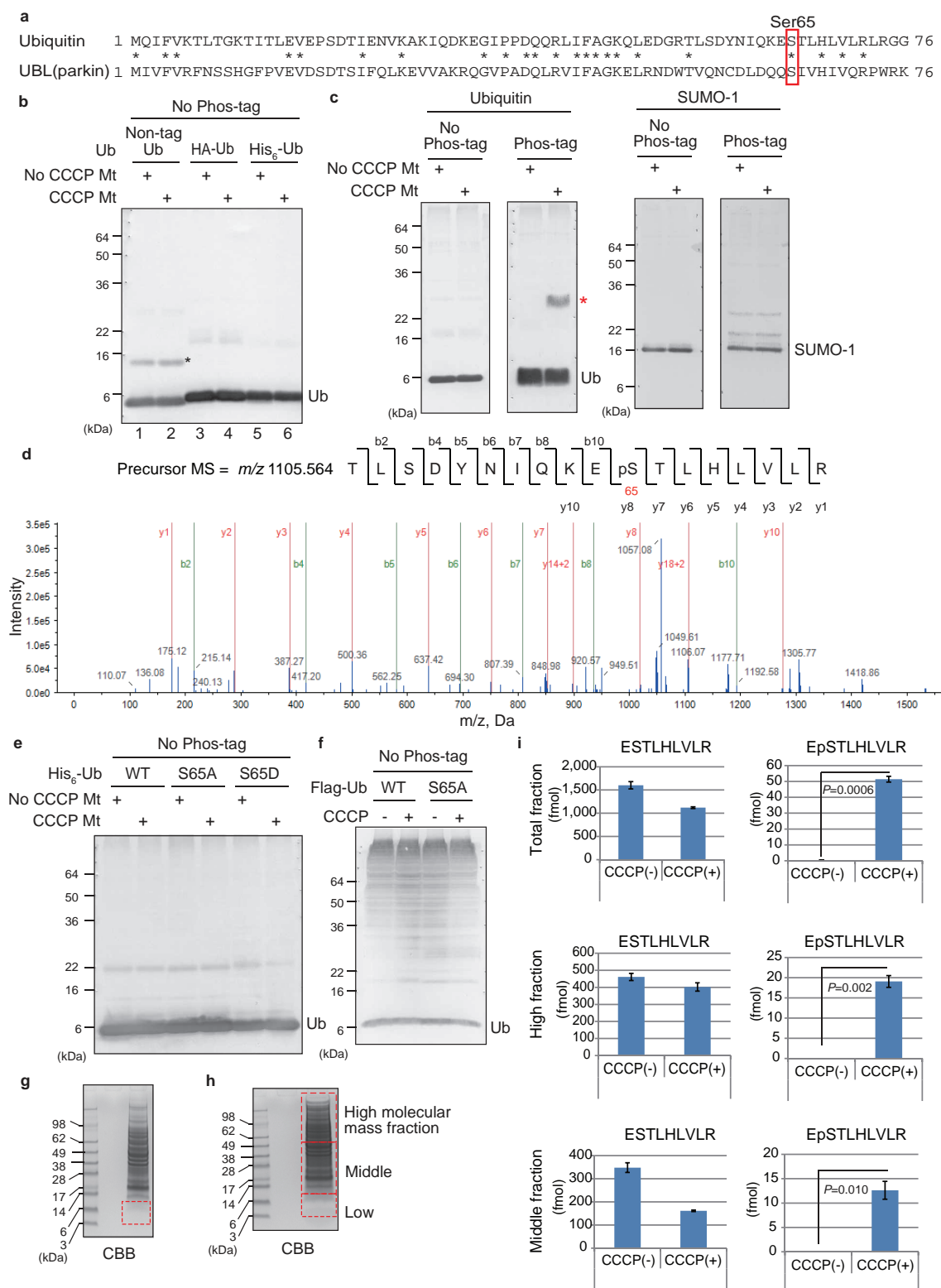
**Yeast analysis.** To confirm the dependency of parkin activation on ubiquitin phosphorylation, we first sought to replace genomic ubiquitin with a phosphorylation-deficient mutant in mammalian cells. However, this type of experiment is challenging

because ubiquitin is encoded at four loci as fourteen copies in the human genome, and synthesized as a fusion protein with essential ribosomal subunits. As an alternative, we used a yeast system in which all the genomic-ubiquitin genes are disrupted, the fused ribosomal proteins (L40 and S31) are complemented, and ubiquitin is expressed from a plasmid<sup>25</sup>. Yeast culture media were prepared according to standard protocols. The Y1168 and Y1169 yeast strains, which express only wild-type ubiquitin and ubiquitin (Ser65Ala) respectively, were created as follows. First, SUB328 (MATa *lys2-801 leu2-3,2-112 ura3-52 his3- $\Delta$ 200 trp1-1 ubi1- $\Delta$ 1::TRP1 ubi2- $\Delta$ 2::ura3 ubi3- $\Delta$ ub-2 ubi4- $\Delta$ 2::LEU2* (pUB146) (pUB100)), which expresses ubiquitin from a galactose-inducible promoter in pUB146 (ref. 25), was transformed with pRS317-Ub and pRS317-Ub(Ser65Ala). pRS317-Ub and pRS317-Ub(Ser65Ala) are LYS2-marked plasmids expressing wild-type or Ser65Ala yeast ubiquitin from a constitutive GPD (glyceraldehyde-3-phosphate dehydrogenase) promoter. The URA3-marked pUB146 plasmid was eliminated by streaking on plates containing 0.5 mg ml<sup>-1</sup> 5-fluoroorotic acid (Wako). The genotype of each strain is as follows: Y1168 (MATa *lys2-801 leu2-3,2-112 ura3-52 his3- $\Delta$ 200 trp1-1 ubi1- $\Delta$ 1::TRP1 ubi2- $\Delta$ 2::ura3 ubi3- $\Delta$ ub-2 ubi4- $\Delta$ 2::LEU2* (pRS317-Ub) (pUB100)) and Y1169 (MATa *lys2-801 leu2-3,2-112 ura3-52 his3- $\Delta$ 200 trp1-1 ubi1- $\Delta$ 1::TRP1 ubi2- $\Delta$ 2::ura3 ubi3- $\Delta$ ub-2 ubi4- $\Delta$ 2::LEU2* (pRS317-Ub(Ser65Ala)) (pUB100)).

To express human parkin in yeast, we cloned the parkin gene into a pADGEW plasmid (multi, TRP1). For genomic integration of human parkin, we created a plasmid (pBS-hphMX4-parkin) with a cassette encoding the hygromycin resistance gene, *hphMX4*, and human parkin tagged at the N terminus with GFP and a 3 $\times$ FLAG epitope tag, and with parkin flanked by the *ADH1* promoter and *CYC1* terminator. The gene cassette amplified from pBS-hphMX4-parkin was introduced into the Y1168 and Y1169 yeast strains described above, and hygromycin-resistant transformants were selected. To localize human PINK1 on the mitochondrial surface in yeast, we fused the first 40 N-terminal residues from a yeast mitochondrial outer membrane protein (Om45) to the N terminus of PINK1 lacking the first 120 N-terminal amino acids (OM-PINK1); note that full-length human PINK1 was unstable in yeast cells. The 40 N-terminal residues of Om45 encompass a transmembrane segment that is known to function as a mitochondrial outer membrane-targeting signal. The DNA fragment encoding OM-PINK1 with a C-terminal 3 $\times$ FLAG-tag was cloned into a URA3-CEN plasmid, p416GAL1. The resulting plasmid, p416GAL1-OM-PINK1-FLAG was introduced into the parkin-expressing yeast strains described above. As the expression of OM-PINK1 is under the control of the *GAL1* promoter, the yeast cells were cultivated in galactose-containing media. Whole-cell lysates were subjected to immunoblot analysis. PINK1 and parkin expression were confirmed by immunoblot analysis with an anti-parkin antibody Prk8 (Cell Signaling Technology) or an anti-PINK1 antibody D8G3 (Cell Signaling Technology).

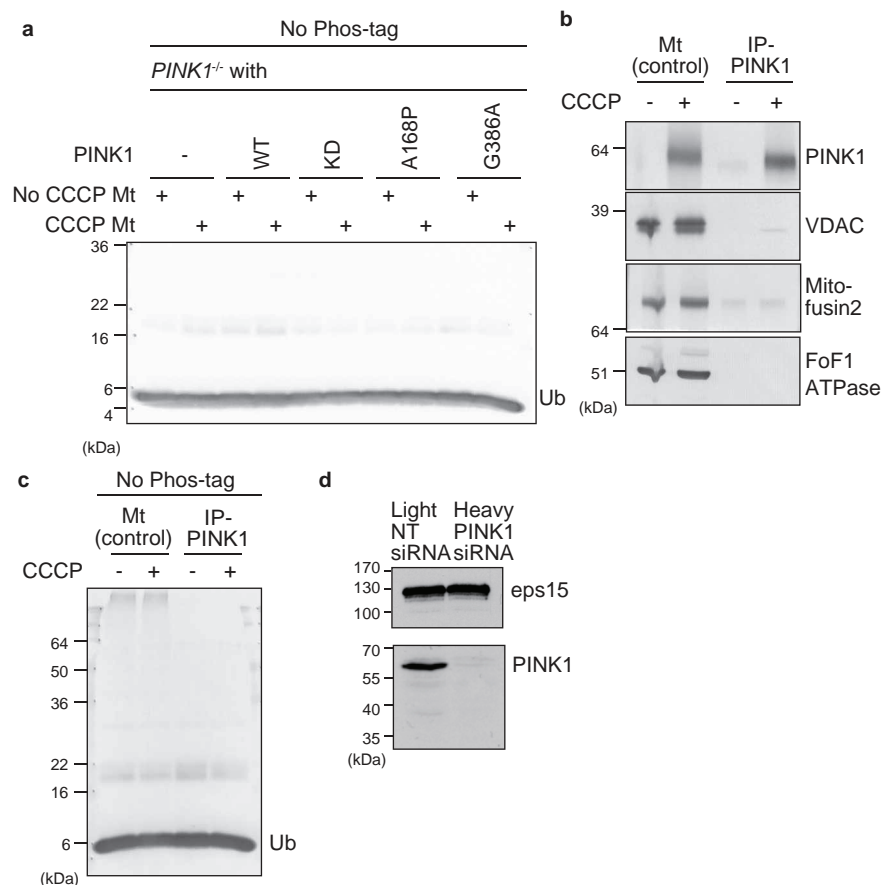
**Computational modelling of fully activated parkin.** The two proposed models for fully activated parkin with structural changes in the RING2 domain incorporating either the phosphorylated Ubl domain or phosphorylated ubiquitin were constructed using the protein docking procedure as follows. We first simulated docking of phosphorylated UBL (structural region from Met 1 to Arg71 from Protein Data Bank (PDB) entry 4k95) or ubiquitin (PDB entry 1ubq) to the structure of inactivate parkin (PDB entry 4k95) in which the RING2 domain was truncated. The RING2 domain was then reassembled in complex with UBL or ubiquitin using protein docking and the full atomic models refined. We used ZDOCK version 3.0.2 to generate 2000 candidate conformation for each protein docking step. During the protein docking procedure, Ser 65 was mutated to glutamate because the ZDOCK program only supports native amino acids. The best model in the first step was manually selected based on the information of a putative phosphopeptide binding site<sup>27</sup> from the structure with high ZDOCK scores. In the second step, we manually selected the final model based on both the ZDOCK score and domain connectivity between REP and RING2 domains. Finally, the Ser65Gln mutation was replaced with phosphorylated Ser using the molecular builder program in Maestro (Schrödinger). Refinement of the final model was performed using the Protein Preparation Wizard workflow in Maestro (Schrödinger).





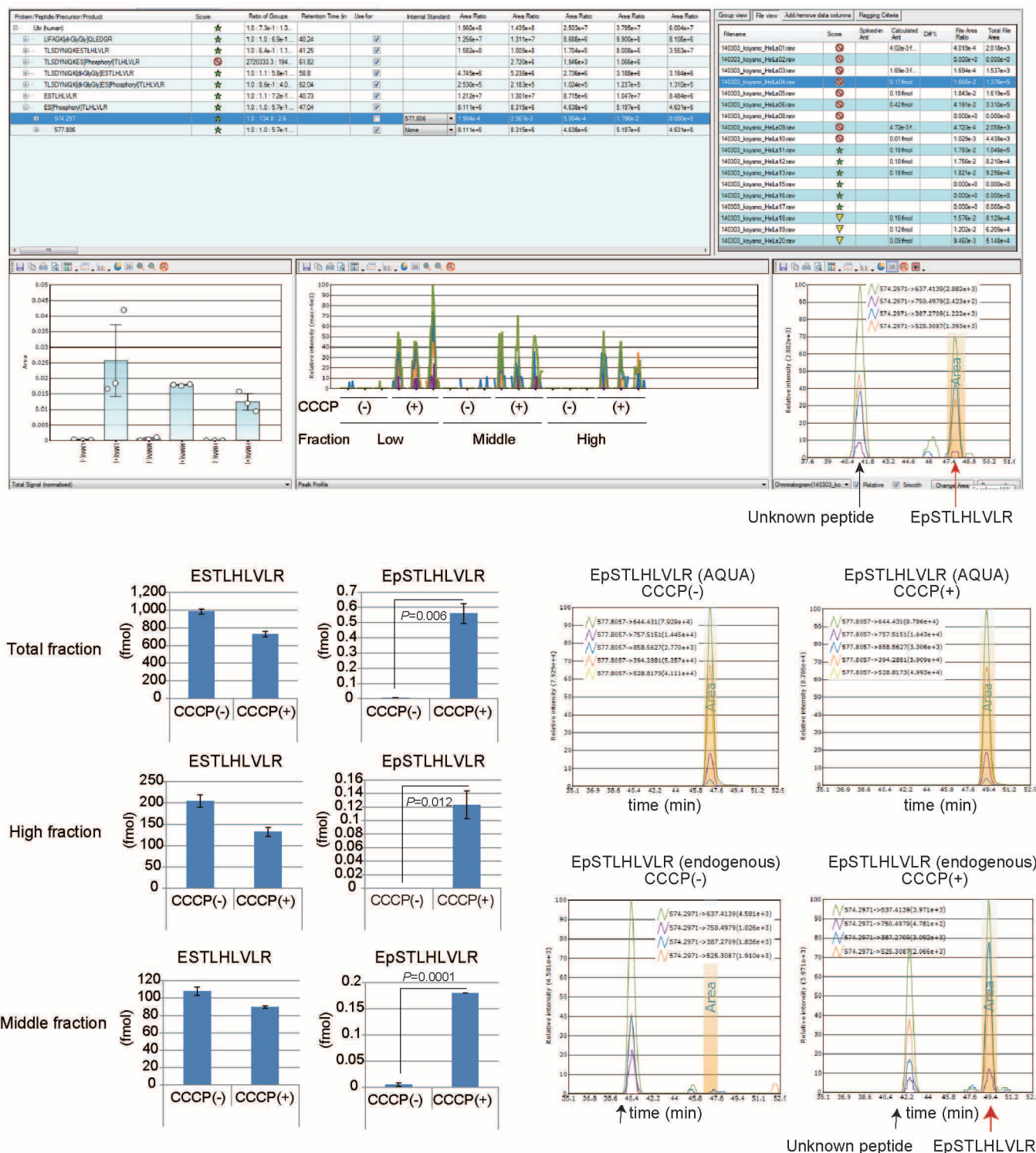
**Extended Data Figure 1 | Ubiquitin Ser 65 is phosphorylated by depolarized mitochondria.** **a**, Sequence comparison between ubiquitin and the parkin UBL domain. Grey asterisks, identical amino acid residues. **b**, When the same samples as Fig. 1b were subjected to non-Phos-tag PAGE, no retarded-mobility form of ubiquitin was observed. Black asterisk, cross-reacting band. **c**, Recombinant SUMO-1 was incubated with depolarized mitochondria prepared from PINK1-expressing HeLa cells *in vitro*. Red asterisk, phosphorylated ubiquitin as a positive control under the same experimental conditions. **d**, Mass spectrometric (MS) analysis of the tryptic phosphopeptide from ubiquitin obtained in Fig. 1b. **e**, **f**, When the same samples as in Fig. 1c (**e**) and Fig. 1d (**f**) were subjected to non Phos-tag PAGE, no retarded-mobility

form of ubiquitin was observed. **g**, The portion of the gel corresponding to the low-molecular-weight fraction of whole-cell extracts used for MS analysis of Fig. 1e. **h**, Whole-cell lysates of PINK1-expressing HeLa cells  $\pm$  CCCP treatment were subjected to SDS-PAGE and the high- (>55 kDa), middle- (14–55 k) and low- (<14 k) molecular-weight fractions were collected for MS analysis of Extended Data Fig. 1i. **i**, Using AQUA peptides as standards, the absolute quantities of the ESTLHLVLR and EpSTLHLVLR peptides in the high, middle and total fractions were determined in three experiments. Error bars represent mean  $\pm$  s.e.m. and statistical significance was calculated using a one-tailed paired *t*-test.



**Extended Data Figure 2 | PINK1 is essential for ubiquitin phosphorylation.**  
**a**, Results of the control experiments for Fig. 2a. The same samples as Fig. 2a were subjected to non-Phos-tag PAGE to confirm that no retarded-mobility form of ubiquitin was observed. **b**, PINK1 was immunoprecipitated and its purity was confirmed by immunoblot analysis with antibodies against other mitochondrial membrane proteins. **c**, Results of the control experiments for

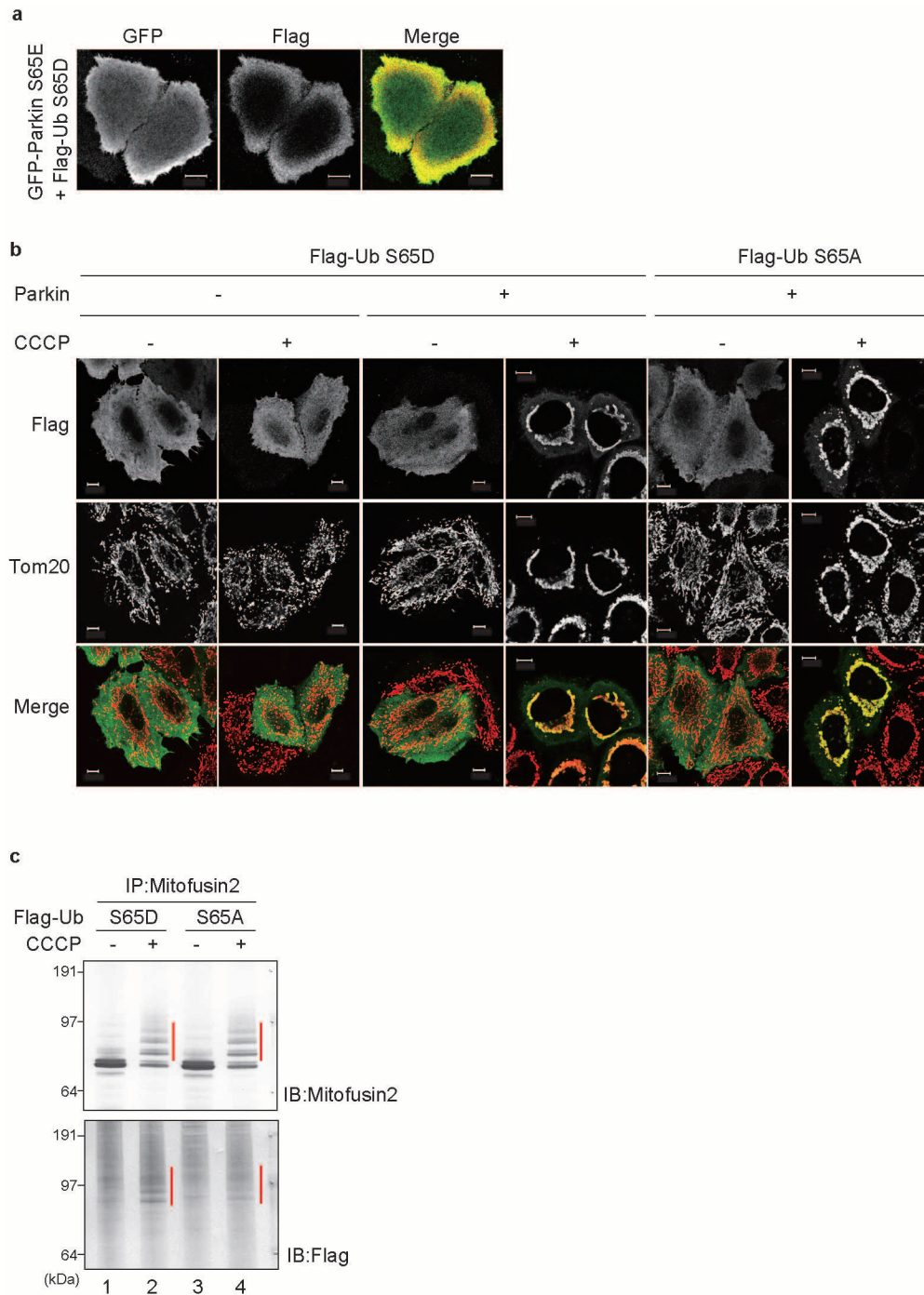
Fig. 2b. The same samples as Fig. 2b were subjected to non Phos-tag PAGE to confirm that no retarded-mobility form of ubiquitin was observed. **d**, Immunoblot analysis to confirm PINK1 knockdown. HeLa cells grown in light or heavy SILAC media were treated with non-target (NT) or *PINK1* siRNA, respectively, and were subjected to immunoblot analysis with an anti-PINK1 antibody.



**Extended Data Figure 3 | Mass-spectrometry-based absolute quantification of phosphorylated ubiquitin in intact HeLa cells.** Whole-cell lysates of intact HeLa cells  $\pm$  CCCP treatment were subjected to SDS-PAGE, and the high- (>55 k), middle- (14–55 k), or low- (<14 k) molecular-weight fractions were collected. The absolute quantities of the ESTLHLVLR and EpSTLHLVLR peptides in the high, middle and total fractions were then determined using AQUA peptides as standards in three experiments. Error bars represent mean

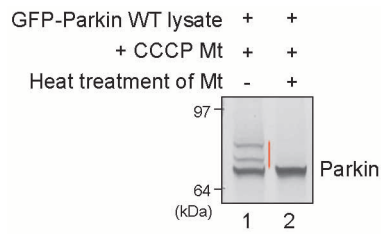
$\pm$  s.e.m. and statistical significance was calculated using a one-tailed paired *t*-test. Because the signal intensity is low, the original data and chromatogram of EpSTLHLVLR peptide are also shown. The light signal (derived from endogenous EpSTLHLVLR peptide), which shows a similar ionized pattern and the identical elution time as the heavy signal (derived from internal AQUA EpSTLHLVLR control peptide), was observed only in CCCP-pretreated cells (red arrows).





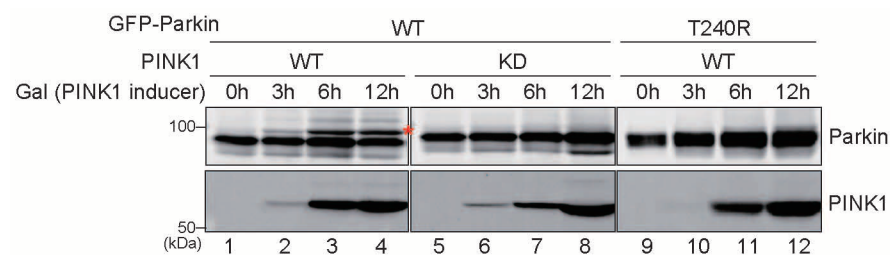
**Extended Data Figure 4 | Subcellular localization of phosphomimetic or phosphorylation-deficient ubiquitin under various experimental conditions.** **a**, Phosphomimetic ubiquitin(Ser65Asp) stimulates the E3 activity of parkin in the absence of mitochondrial localization. Example figures indicative of cytoplasmic localization of both phosphomimetic parkin and ubiquitin are shown. Mitochondrial localization was 0% of 100 cells in two independent experiments. Bars indicate 10  $\mu$ m unless otherwise specified.

**b**, Mitochondrial accumulation of phosphomimetic or phosphorylation-deficient ubiquitin depending on parkin and CCCP treatment (3 h). **c**, Phosphomimetic ubiquitin(Ser65Asp) is conjugated to the parkin substrate mitofusin-2. Red bars indicate ubiquitination of mitofusin-2 with phosphomimetic Flag-ubiquitin (lane 2) or phosphorylation-deficient Flag-ubiquitin (lane 4).



**Extended Data Figure 5 | PINK1 was inactivated following heat treatment.**

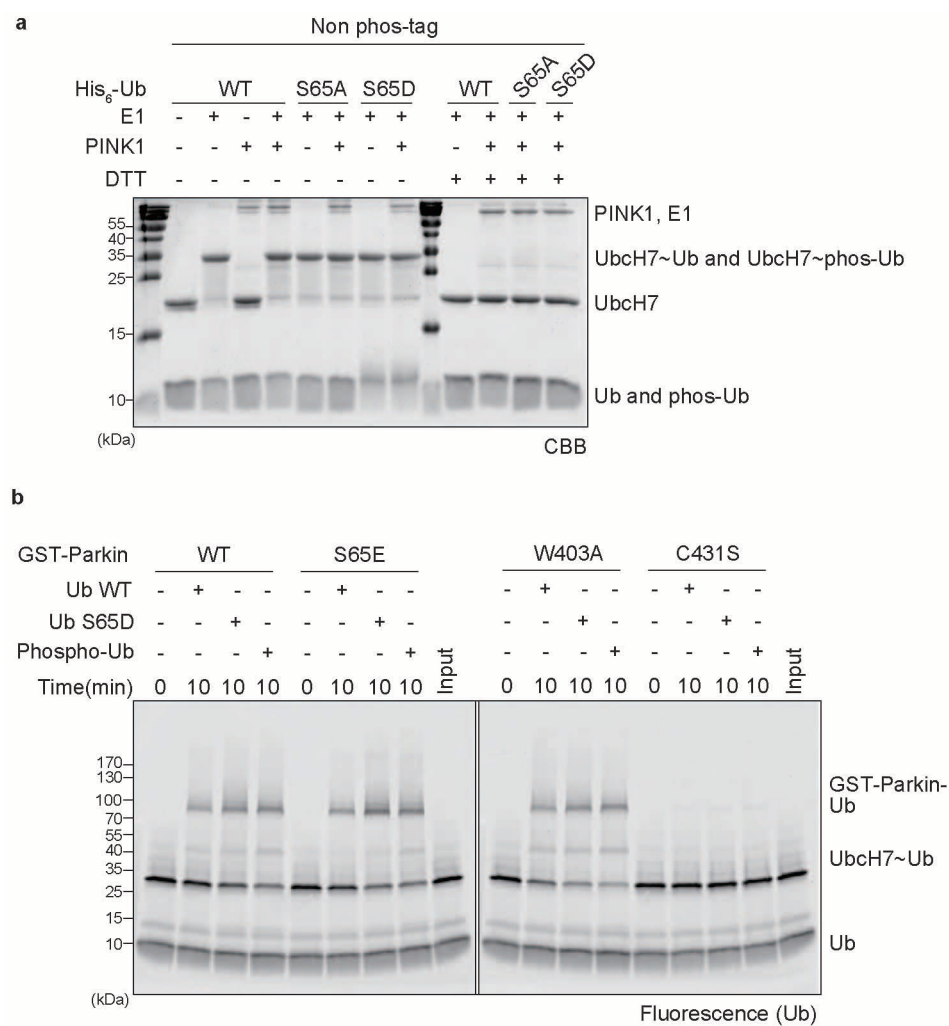
Autoubiquitination of GFP-parkin in cell extracts was reconstituted when incubated with depolarized mitochondria (lane 1), whereas heat pre-treatment of the mitochondria abolished autoubiquitination (lane 2). Because the parkin-activating potency of depolarized mitochondria was lost by heat treatment (at 90 °C for 10 min), mitochondrial proteins including PINK1 were likely inactivated by the aforementioned conditions.



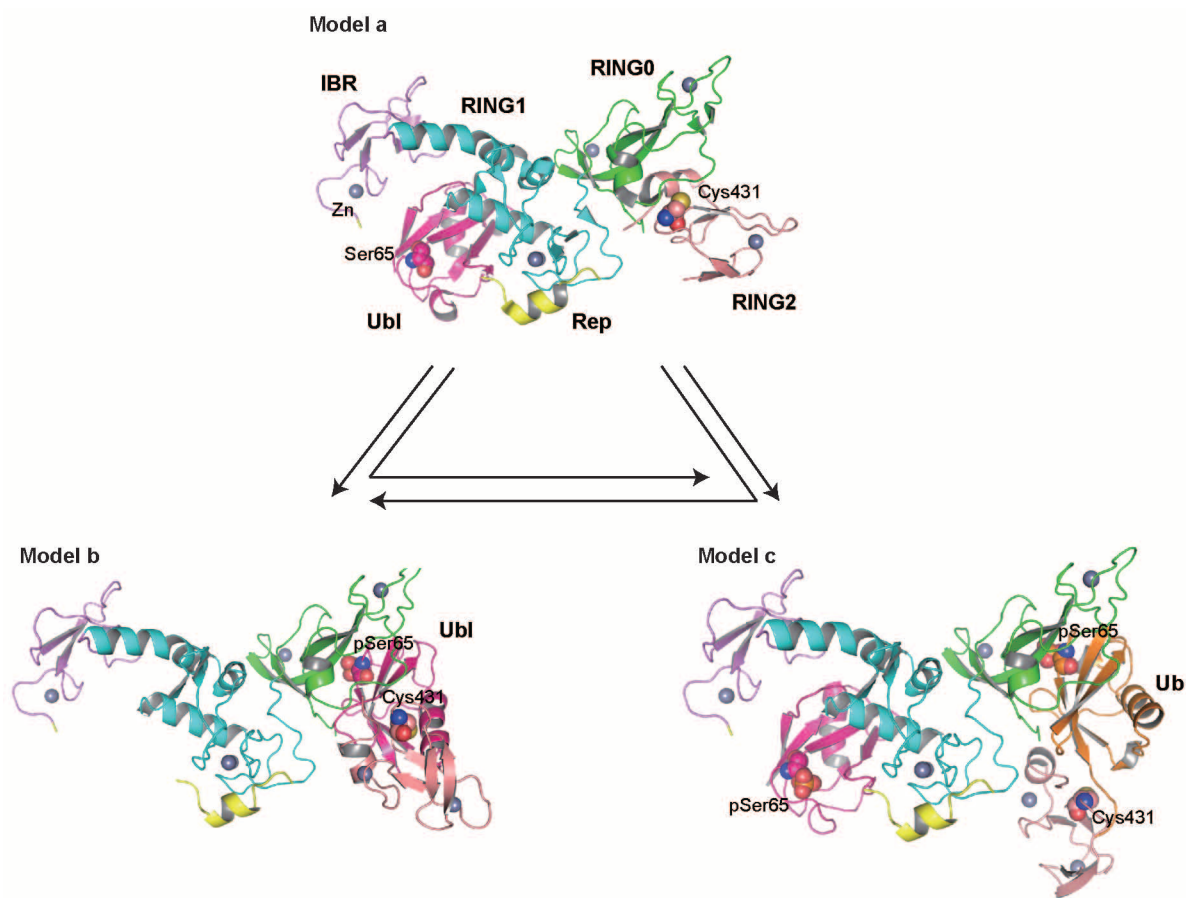
**Extended Data Figure 6 | Reconstitution of parkin activation by PINK1 in yeast cells.** A plasmid encoding GFP-parkin or a pathogenic mutant GFP-parkin(Thr240Arg) under the control of a constitutive promoter and a plasmid encoding OM-PINK1 (40 amino acids at the N terminus of yeast mitochondrial outer-membrane protein Om45 were fused to the N terminus of PINK1 lacking the genuine 120 amino acids at the N terminus) under the control of a galactose (Gal)-inducible promoter were co-transfected into yeast

cells. The yeast cell lysates were collected and subjected to immunoblot analysis with an anti-parkin or anti-PINK1 antibody. As is the case with mammalian cells, a putative autoubiquitination signal was observed with GFP-parkin in accordance with PINK1 mitochondrial accumulation (lanes 1–4), whereas the PINK1 kinase dead (KD) mutation (lanes 5–8) or pathogenic parkin Thr240Arg mutation (lanes 9–12) hindered it.





**Extended Data Figure 7 | *In vitro* charging and discharging assay of UBCH7.** **a**, Charging assay of UBCH7 using both phosphorylated and non-phosphorylated ubiquitin. Samples prepared as in Fig. 4f were resolved on SDS-PAGE using standard (non-Phos-tag) gel. **b**, Discharging assay of UBCH7 using wild-type, phosphomimetic and phosphorylated ubiquitin in the presence of parkin. His-UBCH7 was charged with N-terminal Fluorescein-labelled ubiquitin, and then purified His-UBCH7~FluoUb (1  $\mu$ M) was incubated with wild-type or mutant GST-parkin (0.5  $\mu$ M) in the presence of 10  $\mu$ M His-Ub wild-type, Ser65Asp or phospho-HisUb. Signal derived from Fluorescein-labelled ubiquitin was imaged using the fluorescent scanner. Fluorescence strength of His-UBCH7~FluoUb bands was quantified in three experiments, and were subjected to statistical analysis as shown in Fig. 4g.



#### Extended Data Figure 8 | Structural modelling of the fully activated parkin.

Based on the structure of the inactive, autoinhibited parkin (Model a), two possible models of fully activated parkin with higher accessibility to catalytic cysteine (Cys 431 in the RING2 domain) were generated. Binding of the

phosphorylated UBL domain (Model b) or phosphorylated ubiquitin (Model c) to the RING0 domain supports a structural change in the RING2 domain. It is possible that Model b is an intermediate for Model c, or vice versa.

**Extended Data Table 1 | A list of the ubiquitin peptides used in a parallel reaction monitoring assay to quantify phosphorylated ubiquitin**

Peptide sequence	Precursor m/z (charge state)	Product ions for PRM
ESTLHLVLR	534.314 (+2)	$y_3^+$ , $y_4^+$ , $y_5^+$ , $y_6^+$ , $y_7^+$
ESTLHLVL [Heavy] R	537.823 (+2)	$y_3^+$ , $y_4^+$ , $y_5^+$ , $y_6^+$ , $y_7^+$
EpSTLHLVLR	574.297 (+2)	$y_3^+$ , $y_4^+$ , $y_5^+$ , $y_6^+$ , $y_7^+$
EpSTLHLVL [Heavy] R	577.806 (+2)	$y_3^+$ , $y_4^+$ , $y_5^+$ , $y_6^+$ , $y_7^+$

For each peptide, the optimal precursor ion and product ions are shown.



# Oncogene-like induction of cellular invasion from centrosome amplification

Susana A. Godinho<sup>1,2†</sup>, Remigio Picone<sup>1,2</sup>, Mithila Burute<sup>3,4,5</sup>, Regina Dagher<sup>1,2</sup>, Ying Su<sup>6</sup>, Cheuk T. Leung<sup>2†</sup>, Kornelia Polyak<sup>6</sup>, Joan S. Brugge<sup>2</sup>, Manuel Théry<sup>3,4</sup> & David Pellman<sup>1,2</sup>

**Centrosome amplification has long been recognized as a feature of human tumours; however, its role in tumorigenesis remains unclear<sup>1</sup>. Centrosome amplification is poorly tolerated by non-transformed cells and, in the absence of selection, extra centrosomes are spontaneously lost<sup>2</sup>. Thus, the high frequency of centrosome amplification, particularly in more aggressive tumours<sup>3</sup>, raises the possibility that extra centrosomes could, in some contexts, confer advantageous characteristics that promote tumour progression. Using a three-dimensional model system and other approaches to culture human mammary epithelial cells, we find that centrosome amplification triggers cell invasion. This invasive behaviour is similar to that induced by overexpression of the breast cancer oncogene *ERBB2* (ref. 4) and indeed enhances invasiveness triggered by *ERBB2*. Our data indicate that, through increased centrosomal microtubule nucleation, centrosome amplification increases Rac1 activity, which disrupts normal cell–cell adhesion and promotes invasion. These findings demonstrate that centrosome amplification, a structural alteration of the cytoskeleton, can promote features of malignant transformation.**

The centrosome is the major microtubule-organizing centre in mammalian cells and comprises of a pair of centrioles surrounded by the pericentriolar material<sup>5</sup>. Centrosome abnormalities, usually increased numbers, are common in human tumours<sup>1</sup> and have been positively associated with advanced tumour grade and metastasis<sup>3</sup>, suggesting a possible role in tumour progression. This is somewhat surprising given the well-documented deleterious effects of centrosome amplification on cell proliferation<sup>6</sup>; such amplification can be lethal if it compromises the ability of cells to organize multiple centrosomes to generate pseudo-bipolar spindles<sup>2</sup>. These seemingly paradoxical observations suggest that centrosome amplification might enhance other aspects of tumorigenesis.

We have developed orthogonal approaches to generate genetically comparable cells that do or do not carry extra centrosomes<sup>2</sup>. Here we adapt these methods to determine how centrosome amplification influences epithelial organoid integrity, making use of the well characterized three dimensional (3D) culture model for MCF10A cells, a non-transformed human mammary epithelial cell line. This model recapitulates many aspects of breast glandular architecture<sup>7</sup>.

We engineered MCF10A cells to enable the inducible overexpression of Polo-like kinase 4 (PLK4), an essential regulator of centrosome duplication, whose overexpression induces supernumerary centrosomes<sup>8,9</sup>. As a negative control, we transiently overexpressed a truncated form of PLK4 (PLK4<sup>1–608</sup>) that retains kinase activity but does not induce centrosome amplification<sup>10</sup>. As expected, transient induction of PLK4, but not of PLK4<sup>1–608</sup>, led to centrosome amplification (Fig. 1a and Extended Data Fig. 1). Strikingly, centrosome amplification induced by PLK4 resulted in the formation of invasive protrusions, cytoplasmic extensions that invade the surrounding matrix (Fig. 1b and Extended Data Fig. 1f, g). Expression of centrin1–GFP to visualize the centrioles revealed that virtually all cells with invasive protrusions exhibited centrosome

amplification (Fig. 1c). An independent approach, using an organotypic culture system to assay for fibroblast-led collective migration, confirmed that centrosome amplification promotes invasion, both of MCF10A cells and non-transformed keratinocytes (HaCaTs) (Fig. 1d and Extended Data Fig. 1h).

Cytokinesis failure was induced in MCF10A cells using dihydrocytochalasin B (DCB) to generate centrosome amplification without PLK4 overexpression. Newly generated tetraploid cells, with doubled centrosome content, were isolated by fluorescence-activated cell sorting (FACS). A control population of tetraploid cells in which extra centrosomes were spontaneously lost were generated, as previously described<sup>2</sup> (evolved tetraploids, 4N.evo, Extended Data Fig. 2a–e). Tetraploid cells with extra centrosomes were invasive in 3D cultures, whereas 4N.evo cells were not (Fig. 1e). PLK4 overexpression in 4N.evo cells induced centrosome amplification accompanied by invasive protrusions, demonstrating that 4N.evo cells still retained the ability to become invasive (Extended Data Fig. 2g, h).

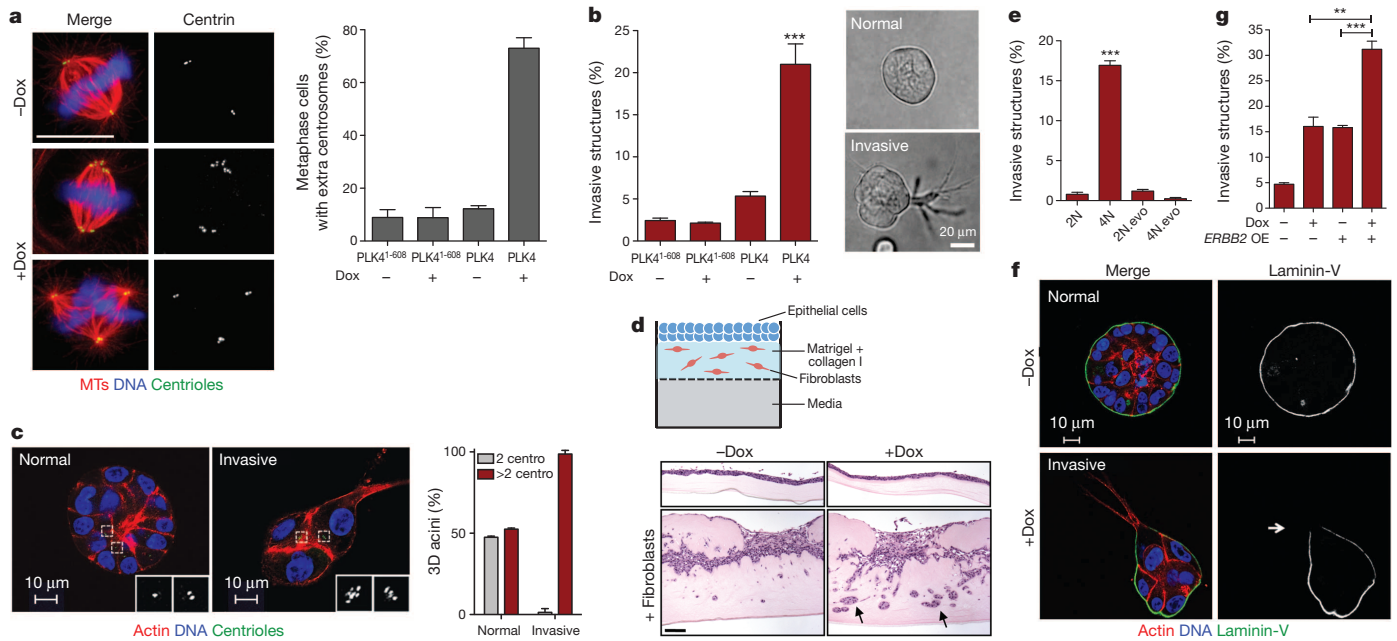
Invasive protrusions are accompanied by the degradation of laminin-V (Fig. 1f) and collagen-I (Extended Data Fig. 1i), contain actin and microtubules (Extended Data Fig. 3a) and are surrounded by the extracellular matrix component fibronectin (Extended Data Fig. 3b). Consistent with centrosome amplification promoting matrix degradation, the invasive phenotype was partially suppressed by inhibition of metalloproteinases using marimastat (Extended Data Fig. 3c). Live-cell imaging showed that protrusions are highly dynamic, constantly extending and retracting (Supplementary Videos 1 and 2), which may partially explain why only a fraction of acini with extra centrosomes exhibits invasive protrusions at a given time (Fig. 1c). The formation of an initial protrusion provided a track for the collective migration of multiple cells out of the acinus and into the surrounding matrix (Extended Data Fig. 3d, e and Supplementary Video 3). This type of collective invasion resembles what has been observed in tumours *in vivo*. Indeed, many solid tumours typically exhibit collective invasion, which often involves the degradation of the extracellular matrix<sup>11</sup>.

Invasion induced by centrosome amplification strongly resembled that induced by a bona fide breast cancer oncogene, *ERBB2* (ref. 4) (Extended Data Fig. 4). As in *ERBB2* tumours<sup>12</sup>, cells with extra centrosomes retained the expression of E-cadherin (Extended Data Fig. 3f), suggesting that mechanisms other than a classical epithelial-to-mesenchymal transition account for the invasive phenotype. Importantly, when combined with *ERBB2* overexpression, centrosome amplification enhanced the frequency of invasive acini (Fig. 1g).

The induction of invasive protrusions by extra centrosomes could be an indirect consequence of aneuploidy that results from chromosome missegregation<sup>2,13</sup>. To directly address this possibility, we depleted MCAK (also known as KIF2C), a kinesin important for chromosome segregation during mitosis<sup>14</sup>, to induce a comparable degree of aneuploidy before (48 h after PLK4 induction) and after 4 days in 3D culture (Fig. 2a, b).

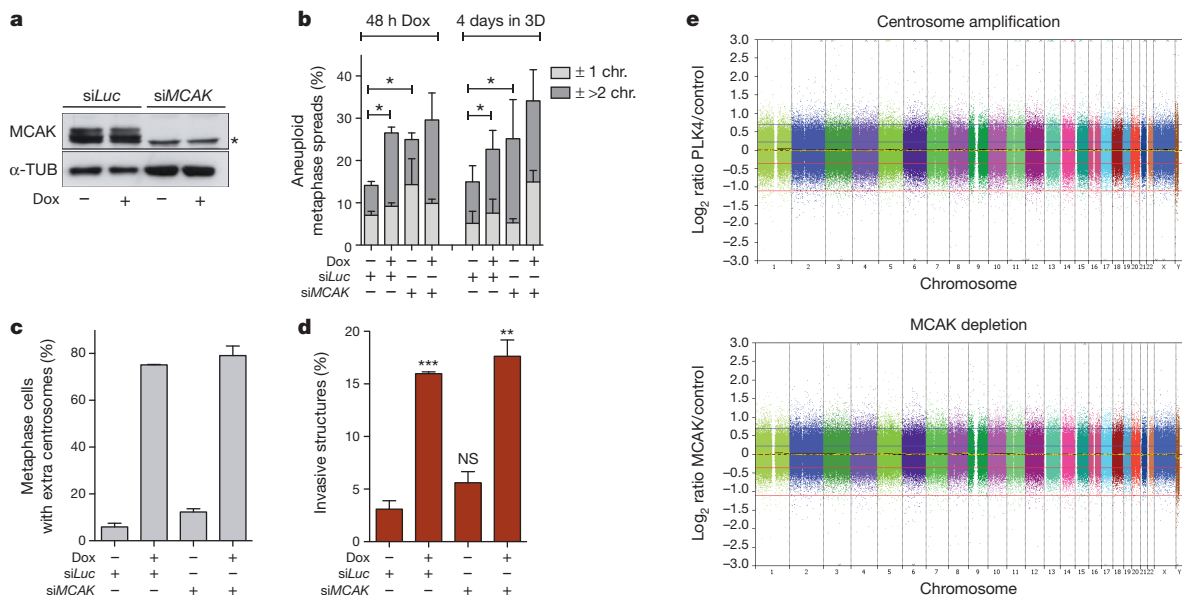
<sup>1</sup>Howard Hughes Medical Institute, Department of Pediatric Oncology, Dana-Farber Cancer Institute and Pediatric Hematology/Oncology, Children's Hospital, Boston, Massachusetts 02115, USA.

<sup>2</sup>Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>3</sup>Institut de Recherche en Technologie et Science pour le Vivant, UMR5168 CEA/UJF/INRA/CNRS, Grenoble, France. <sup>4</sup>Hôpital Saint Louis, Institut Universitaire d'Hématologie, U1160 INSERM/AP-HP/Université Paris Diderot, Paris 75010, France. <sup>5</sup>CYT00 SA, Grenoble 38054, France. <sup>6</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>†</sup>Present addresses: Barts Cancer Institute, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK (S.A.G.); Department of Pharmacology, University of Minnesota, Minneapolis, Minnesota 55455, USA (C.T.L.).



**Figure 1 | Invasive behaviour of epithelial cells triggered by centrosome amplification.** **a**, Cells stained for microtubules ( $\alpha$ -tubulin, red), centrioles (centrin2, green) and DNA (Hoechst, blue). Scale bar, 10  $\mu$ m. Histogram shows fraction of cells with centrosome amplification. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **b**, Fraction of invasive acini in 3D cultures and representative images of normal acini and an acinus with invasive protrusions. Scale bar, 10  $\mu$ m. Error bars represent mean  $\pm$  s.e. from 4 independent experiments. **c**, Cells stained for F-actin (red), centrioles (centrin1-GFP, green, inset white) and DNA (blue). Scale bar, 10  $\mu$ m. Histogram shows fraction of acini with centrosome amplification after PLK4 overexpression. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **d**, Scheme of the organotypic culture model used to assess

invasion (top). Images show haematoxylin and eosin (H&E) staining of sections of MCF10A cells plated on the organotypic model, with and without fibroblasts (black arrows indicate highly invasive areas). Percentage of invasion (mean  $\pm$  s.d.): -Dox =  $11.7 \pm 0.83$ ; +Dox =  $26.1 \pm 6.5$ . Scale bar, 100  $\mu$ m. **e**, Fraction of invasive acini in tetraploids. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **f**, Acini stained for laminin-V (green), F-actin (red) and DNA (blue). White arrow indicates laminin-V degradation. Scale bar, 10  $\mu$ m. **g**, Fraction of invasive acini in cells that overexpress *ERBB2*, with or without centrosome amplification. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. All *P* values were derived from unpaired two-tailed Student's *t*-test (\*\**P* < 0.005; \*\*\**P* < 0.0005). OE, overexpression.



**Figure 2 | The induction of aneuploidy does not generate invasive acini.** **a**, Western blot showing depletion of MCAK by siRNA (48 h). The asterisk marks a non-specific band. **b**, Quantification of chromosome number, shown as the percentage deviation from the mode, in cells before adding to 3D cultures (48 h) and after 3D cultures (~100 spreads were scored per condition). Significance was determined by chi-squared testing to compare the frequencies of nominal variables (\**P* < 0.05). siLuc, si luciferase control. **c**, Fraction of cells

with centrosome amplification. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **d**, Fraction of invasive acini. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. *P* value derived from unpaired two-tailed Student's *t*-test (\*\**P* < 0.005, \*\*\**P* < 0.0005; NS, not significant). **e**, SNP analysis of cells with extra centrosomes and depleted of MCAK. Shown are log<sub>2</sub> copy number ratios of the indicated samples relative to their controls.

Cells depleted of MCAK neither increased centrosome number nor exhibited a significant increase in invasive protrusions (Fig. 2c, d). Similarly, aneuploidy, generated by inhibition of the spindle assembly checkpoint kinase MPS1 (using the inhibitor reversine)<sup>15</sup> also failed to induce invasion (Extended Data Fig. 5a–c). In addition, although 4N.ev cells exhibit substantial aneuploidy, they did not form invasive acini (Fig. 1e and Extended Data Fig. 2f). Finally, single-nucleotide polymorphism analysis (SNP-arrays) demonstrated that neither cells with extra centrosomes nor depletion of MCAK accumulate a recurrent aneuploidy after 4 days in 3D culture (Fig. 2e and Extended Data Fig. 5d). Thus, aneuploidy per se is not responsible for the invasive behaviour.

Centrosome amplification also leads to altered cilia signalling<sup>16</sup> and to increased levels of p53 (ref. 17). However, we found that neither the parental MCF10A cells nor the derivatives with extra centrosomes formed detectable cilia after 4 days in 3D culture. Moreover, depletion of p53 did not alter centrosome amplification mediated invasion (Extended Data Fig. 6a–d). In a transplant model using asymmetrically dividing neuroblasts from *Drosophila melanogaster*, centrosome amplification induces tumours that, interestingly, are capable of metastasis<sup>18</sup>. Centrosome amplification in this system disrupts asymmetric cell division, resulting in stem cell expansion, potentially contributing to tumorigenesis. Because MCF10A cells do not undergo asymmetric cell division, this mechanism does not apply to our results. In the *Drosophila* neuroblasts, centrosome amplification may also disrupt cell polarization, which can trigger tumorigenesis in many systems. Although we cannot exclude effects on cell polarization in our system, we do note that MCF10A cells cannot form tight junctions and do not exhibit polarization of the apical

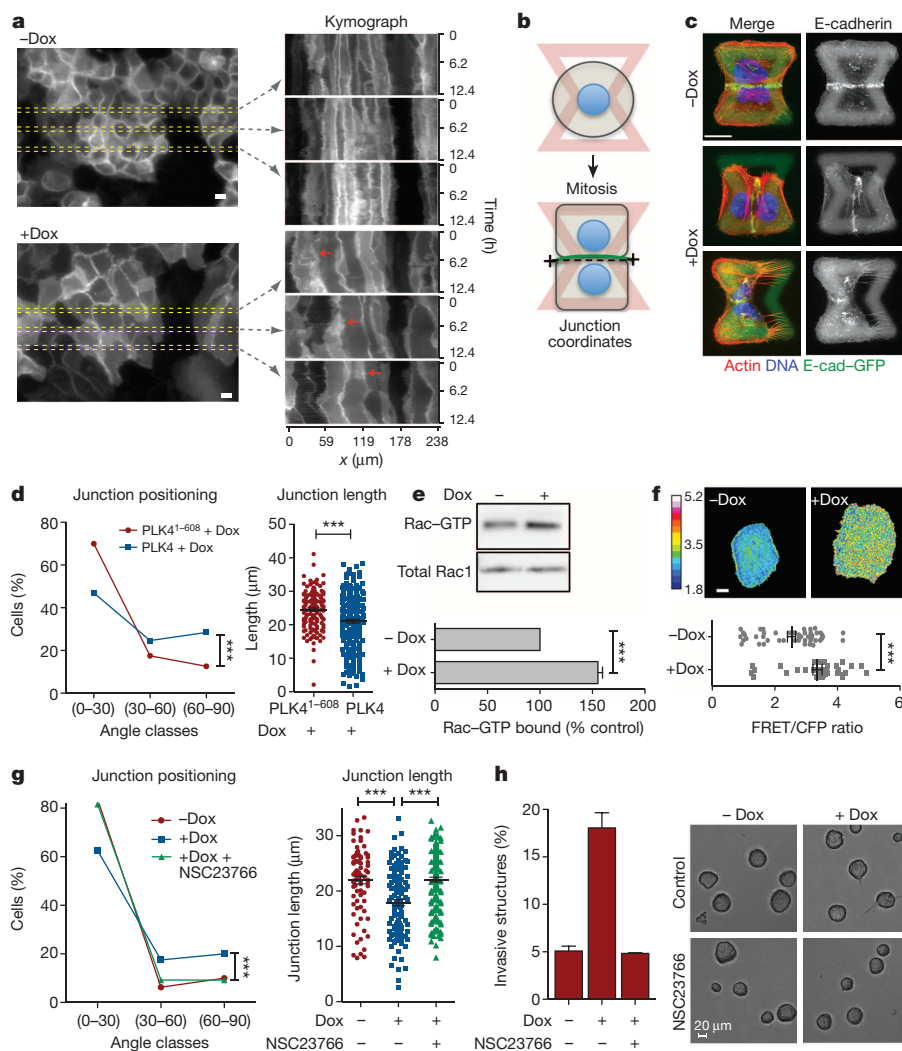
Par3–Par6–aPKC complex<sup>19</sup>. Moreover, centrosome amplification does not impair the ability of MCF10A cells to asymmetrically position centrosomes, the main detectable polarization in these cells (Extended Data Fig. 6e, f).

Insight into why cells with extra centrosomes are invasive in 3D cultures first came from observing the adhesive properties of single cells after cell division. As expected because they are epithelial cells, following mitosis, MCF10As formed cell–cell contacts, and remained as apposed cell pairs. By contrast, cells with extra centrosomes ‘scattered’, resulting in a high fraction of individual cells (Extended Data Fig. 7a, b and Supplementary Videos 4 and 5), a characteristic associated with loss of cell–cell adhesion. Furthermore, live cell imaging in cells with a fluorescent membrane marker revealed that cell–cell contacts are not stable and often overlap in cells with extra centrosomes (Fig. 3a, Supplementary Videos 6 and 7). As a direct measure of cell–cell junction integrity, we plated cells onto fibronectin micropatterns that were specifically designed to promote the formation of adherens junctions at a stereotypical position between the two cells (Fig. 3b)<sup>20</sup>. Consistent with the cell scattering effect, centrosome amplification produced marked defects in cell–cell junction positioning and size (Fig. 3c, d). This effect of centrosome amplification is similar to that described for loss of p120 catenin, whose knockdown weakens cell–cell contacts<sup>20</sup>, although centrosome amplification does not affect p120 levels (Extended Data Fig. 7c).

These phenotypic characteristics induced by centrosome amplification are similar to what has previously been observed upon activation of Rac1 (refs 21, 22), a small GTPase strongly associated with oncogenic signalling and with the induction of invasiveness and metastasis<sup>23,24</sup>.

### Figure 3 | Centrosome amplification disrupts normal cell–cell adhesion because of Rac1 activation.

**a**, Kymograph analysis of cell–cell adhesion in live cells visualized with a plasma membrane marker (Raichu-Rac1, CFP fluorescence). Red arrows mark areas of increased dynamics of cell–cell contacts. **b**, Scheme of the micro-pattern used. **c**, Images of cells (after mitosis) on micro-patterns labelled for E-cadherin (GFP; green), F-actin (red), DNA (blue) and fibronectin micro-pattern (green). Normal sized junction with a normal position (narrow angle) (top); abnormal junction position (middle); smaller junction size (bottom). Scale bar, 10  $\mu$ m. **d**, Distribution of the cell–cell junctions angles, as a measure of junction positioning, (left) and size (right). For –Dox,  $n = 143$ ; +Dox,  $n = 216$ . Error bars represent mean  $\pm$  s.e. **e**, Representative western blot from a pull-down experiment to detect GTP-bound Rac1. Histogram shows quantification from 3 independent pull-down experiments. Error bars represent mean  $\pm$  s.e. **f**, Top shows examples of FRET ratiometric images; colour-coded scale represents the level of Rac1 activation. Scale bar, 10  $\mu$ m. Bottom, levels of active Rac1 measured by FRET. For –Dox,  $n = 50$ ; +Dox,  $n = 38$ . Error bars represent mean  $\pm$  s.e. **g**, Inhibition of Rac1 with NSC23766 in cells plated on micro-patterns. For –Dox,  $n = 80$ ; +Dox,  $n = 119$ ; +Dox + NSC23766,  $n = 162$ . **h**, Fraction of invasive acini after treatment with NSC23766 and images of control and NSC23766-treated acini. Scale bar, 20  $\mu$ m. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. All  $P$  values were derived from unpaired two-tailed Student's  $t$ -test ( $***P < 0.0005$ ).





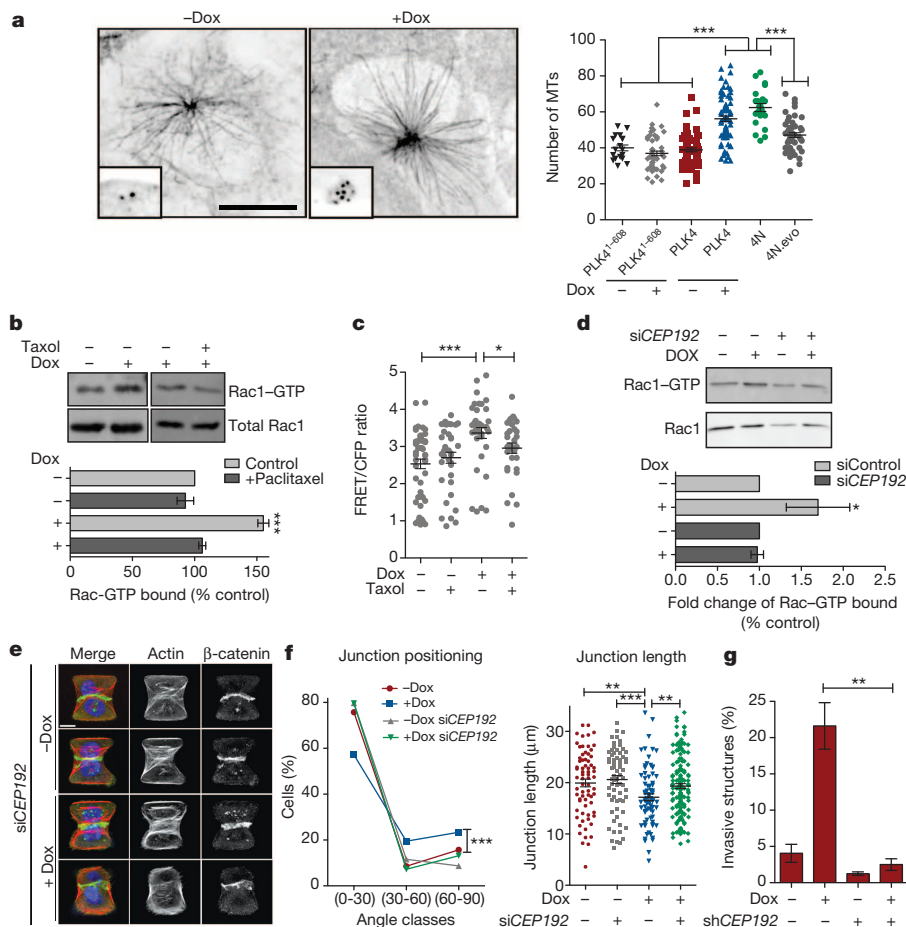
This motivated the hypothesis that extra centrosomes might promote invasive-like effects through inappropriate activation of Rac1. Indeed, we found that centrosome amplification induced a consistent ~1.5-fold Rac1 activation using a biochemical pull-down assay to measure GTP-bound Rac1 in multiple cell lines. Maximal Rac1 activity induced by EGF in MCF10A cells is ~2-fold (Fig. 3e and Extended Data Fig. 7d, e). This was confirmed by monitoring Rac1 activation in single cells using the Raichu-Rac1 fluorescence resonance energy transfer (FRET) biosensor<sup>25</sup> (Fig. 3f and Extended Data Fig. 7f, g). Consistent with the commonly reported antagonism between Rac1 and RhoA, we found that cells with extra centrosomes have decreased active RhoA (Extended Data Fig. 7h). Thus, centrosome amplification in MCF10A cells activates Rac1.

Small molecule Rac1 inhibitor (NSC23766) inhibited Rac1 activation and partially rescued defects in cell–cell adhesion, suggesting that Rac1 activation is responsible for the cell–cell adhesion defect in cells with centrosome amplification, (Fig. 3g and Extended Data Fig. 7i). Similar results were obtained in tetraploid MCF10A cells (Extended Data Fig. 8a, b). Furthermore, treatment with CK-666, an inhibitor of the Arp2/3 complex, an actin nucleator and important downstream target of Rac1 (ref. 23), also partially rescued the defects in cell–cell adhesion induced by centrosome amplification (Extended Data Fig. 8c–e). These findings demonstrate that the cell–cell adhesion defects in cells with extra centrosomes are, to a substantial degree, caused by increased Arp2/3-dependent actin polymerization that occurs downstream of Rac1 signalling. In addition, we found that Rac1 inhibition blocked the formation of invasive acini without impairing the ability of cells to form normal acini (Fig. 3h).

Previous work has established that microtubule polymerization after nocodazole washout induces Rac1 activation. This activation appears to require dynamic microtubules because it is suppressed by the microtubule-stabilizing agent, paclitaxel<sup>26</sup>. We considered the possibility that centrosome amplification induces Rac1 activation through effects on centrosomal

microtubule nucleation. As expected<sup>27</sup>, MCF10A cells with extra centrosomes display elevated levels of centrosomal  $\gamma$ -tubulin (Extended Data Fig. 9a–c) and an increased capacity for microtubule-nucleation (Fig. 4a, note that amplified centrosomes are almost always clustered in interphase). Moreover, paclitaxel blocked the activation of Rac1 in cells with extra centrosomes (Fig. 4b), indicating a similar requirement for dynamic microtubules<sup>26</sup>. These results were independently confirmed when Rac1 activity was monitored by FRET (Fig. 4c). Finally, Rac1 activation measured by FRET was also observed in cells with extra centrosomes deprived of EGF (Extended Data Fig. 9d), indicating that this Rac1 activation is independent of any effects on growth factor signalling. Furthermore, induction of centrosome amplification (at 48 h) does not alter the cell cycle profile (Extended Data Fig. 9e) arguing against cell cycle effects as the cause for Rac1 activation in cells with extra centrosomes.

We next examined the impact of increased centrosomal microtubule nucleation on cell–cell adhesion and the development of invasive structures. To do so, centrosomal microtubule nucleation was moderately decreased by RNAi-mediated knockdown of *CEP192*, which encodes a centrosomal protein required for interphase recruitment of  $\gamma$ -tubulin to the centrosomes<sup>28</sup>. As described<sup>28</sup>, this protocol decreases centrosomal  $\gamma$ -tubulin without affecting centrosome number (Extended Data Fig. 10a, b, d, e). Consistent with our hypothesis, the depletion of CEP192 inhibited Rac1 activation and restored normal cell–cell adhesion among cells with centrosome amplification (Fig. 4d–f). Most notably, partial depletion of CEP192 after short hairpin RNA (shRNA) treatment fully suppressed the invasive phenotype in cells with extra centrosomes without compromising cell viability or centrosome amplification, even after 4 days in 3D cultures (Fig. 4g and Extended Data Fig. 10c, f–h). These results indicate that increased centrosomal microtubule nucleation in cells with extra centrosomes triggers invasion. Although it remains unclear how dynamic microtubules activate Rac1 (ref. 29), our data



**Figure 4 | Effects of centrosome amplification are mediated by increased nucleation of centrosomal microtubules.** **a**, Images of microtubules ( $\alpha$ -tubulin; insets, centrioles) in cells after microtubule re-growth. Graph shows microtubule numbers from the indicated cells. For *PLK4*<sup>1-608</sup> -Dox, *n* = 18; *PLK4*<sup>1-608</sup> +Dox, *n* = 51; *PLK4* -Dox, *n* = 66; *PLK4* +Dox, *n* = 71; 4N, *n* = 22; 4N.evo, *n* = 49. Error bars represent mean  $\pm$  s.e. **b**, Pull-down assay to measure GTP-bound Rac1 after paclitaxel treatment and quantification of the levels of Rac1-GTP. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **c**, FRET ratios for measuring active Rac1 in cells after the indicated treatments. For -Dox, *n* = 50; -Dox + Taxol, *n* = 38; +Dox, *n* = 38; +Dox + Taxol, *n* = 32. Error bars represent mean  $\pm$  s.e. **d**, Pull-down assay to measure GTP-bound Rac1 after CEP192 depletion by siRNA and quantification of the levels of Rac1-GTP. Error bars represent mean  $\pm$  s.e. from 4 independent experiments. **e**, Images of cells depleted of CEP192 on micro-patterns labelled for  $\beta$ -catenin (green), F-actin (red), DNA (blue) and fibronectin micro-pattern (green). Scale bar, 10  $\mu$ m. **f**, Angles and sizes of cell–cell junctions after depletion of CEP192 by siRNA. For control siRNA -Dox, *n* = 71; control siRNA +Dox, *n* = 78; CEP192 siRNA -Dox, *n* = 69; CEP192 siRNA +Dox, *n* = 150. Error bars represent mean  $\pm$  s.e. **g**, Fraction of invasive acini seen after depletion of CEP192 by shRNA. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. For panels a–c, f and g, *P* values were derived from unpaired two-tailed Student's *t*-test (\**P* < 0.05, \*\**P* < 0.005, \*\*\**P* < 0.0005).

indicates that Rac1 activation downstream of microtubule nucleation plays a central role in this phenomenon.

Deregulated Rac1 activity has been implicated in the pathogenesis of many tumour types and is known to drive tumour invasion and metastasis<sup>24</sup>. Previous studies have demonstrated that Rac1 activation in tumours can be regulated by many mechanisms<sup>24</sup>. Our data indicate that centrosome amplification is probably another common mechanism for Rac1 activation during tumorigenesis, and that centrosome amplification may augment other oncogenic signals. In addition to Rac1 activation, additional mechanisms probably contribute to the effects of centrosome amplification. Gene expression analyses of 3D cultures showed that TGF- $\beta$  pathway, involved in metastasis formation and associated with tumour aggressiveness<sup>30</sup>, is strongly upregulated in cells with extra centrosomes (S.A.G. and D.P., unpublished data).

Many studies have noted a positive correlation between centrosome amplification and advanced-stage tumours, recurrence and poor survival<sup>3</sup>, yet the mechanistic basis for this correlation has remained unclear. Here we provide evidence that centrosome amplification can mimic and accentuate the effects of oncogenes in triggering cellular invasion. These findings illustrate the integral relationship between cellular signalling and the cytoskeleton, underscoring the importance of this relationship to tumour progression.

## METHODS SUMMARY

Human mammary epithelial MCF10A cells were maintained at 37 °C with 5% CO<sub>2</sub> atmosphere. PLK4 overexpression to induce centrosome amplification was induced by treatment with 2  $\mu$ g ml<sup>-1</sup> of doxycycline (48 h). Alternatively, centrosome amplification was induced by inhibition of cytokinesis with 4  $\mu$ M of DCB (18 h) and tetraploid cells were isolated by FACS, as previously described<sup>2</sup>. A more detailed description of methods used can be found in the Methods section.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 July 2013; accepted 24 March 2014.

Published online 13 April; corrected online 2 May 2014 (see full-text HTML version for details).

- Zyss, D. & Gergely, F. Centrosome function in cancer: guilty or innocent? *Trends Cell Biol.* **19**, 334–346 (2009).
- Ganem, N. J., Godinho, S. A. & Pellman, D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460**, 278–282 (2009).
- Chan, J. Y. A clinical overview of centrosome amplification in human cancers. *Int. J. Biol. Sci.* **7**, 1122–1144 (2011).
- Levental, K. R. *et al.* Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* **139**, 891–906 (2009).
- Bettencourt-Dias, M. & Glover, D. M. Centrosome biogenesis and function: centrosomes brings new understanding. *Nature Rev. Mol. Cell Biol.* **8**, 451–463 (2007).
- Sluder, G. & Nordberg, J. J. The good, the bad and the ugly: the practical consequences of centrosome amplification. *Curr. Opin. Cell Biol.* **16**, 49–54 (2004).
- Debnath, J. & Brugge, J. S. Modelling glandular epithelial cancers in three-dimensional cultures. *Nature Rev. Cancer* **5**, 675–688 (2005).
- Bettencourt-Dias, M. *et al.* SAK/PLK4 is required for centriole duplication and flagella development. *Curr. Biol.* **15**, 2199–2207 (2005).
- Habedanck, R., Stierhof, Y. D., Wilkinson, C. J. & Nigg, E. A. The Polo kinase Plk4 functions in centriole duplication. *Nature Cell Biol.* **7**, 1140–1146 (2005).
- Guderian, G., Westendorf, J., Uldschmid, A. & Nigg, E. A. Plk4 trans-autophosphorylation regulates centriole number by controlling  $\beta$ TRCP-mediated degradation. *J. Cell Sci.* **123**, 2163–2169 (2010).
- Friedl, P., Locker, J., Sahai, E. & Segall, J. E. Classifying collective cancer cell invasion. *Nature Cell Biol.* **14**, 777–783 (2012).
- Palacios, J. *et al.* Relationship between ERBB2 and E-cadherin expression in human breast cancer. *Virchows Archiv.* **427**, 259–263 (1995).
- Silkworth, W. T., Nardi, I. K., Scholl, L. M. & Cimini, D. Multipolar spindle pole coalescence is a major source of kinetochore mis-attachment and chromosome mis-segregation in cancer cells. *PLoS ONE* **4**, e6564 (2009).
- Maney, T., Hunter, A. W., Wagenbach, M. & Wordeman, L. Mitotic centromere-associated kinesin is important for anaphase chromosome segregation. *J. Cell Biol.* **142**, 787–801 (1998).
- Santaguida, S., Tighe, A., D'Alise, A. M., Taylor, S. S. & Musacchio, A. Dissecting the role of MPS1 in chromosome biorientation and the spindle checkpoint through the small molecule inhibitor reversine. *J. Cell Biol.* **190**, 73–87 (2010).
- Mahjoub, M. R. & Stearns, T. Supernumerary centrosomes nucleate extra cilia and compromise primary cilium signaling. *Curr. Biol.* **22**, 1628–1634 (2012).
- Holland, A. J. *et al.* The autoregulated instability of Polo-like kinase 4 limits centrosome duplication to once per cell cycle. *Genes Dev.* **26**, 2684–2689 (2012).
- Basto, R. *et al.* Centrosome amplification can initiate tumorigenesis in flies. *Cell* **133**, 1032–1042 (2008).
- Fogg, V. C., Liu, C. J. & Margolis, B. Multiple regions of Crumbs3 are required for tight junction formation in MCF10A cells. *J. Cell Sci.* **118**, 2859–2869 (2005).
- Tseng, Q. *et al.* Spatial organization of the extracellular matrix regulates cell-cell junction positioning. *Proc. Natl Acad. Sci. USA* **109**, 1506–1511 (2012).
- Chen, X. & Macara, I. G. Par-3 controls tight junction assembly through the Rac exchange factor Tiam1. *Nature Cell Biol.* **7**, 262–269 (2005).
- Xue, B., Krishnamurthy, K., Allred, D. C. & Muthuswamy, S. K. Loss of Par3 promotes breast cancer metastasis by compromising cell–cell cohesion. *Nature Cell Biol.* **15**, 189–200 (2013).
- Jaffe, A. B. & Hall, A. Rho GTPases: biochemistry and biology. *Annu. Rev. Cell Dev. Biol.* **21**, 247–269 (2005).
- Mack, N. A., Whalley, H. J., Castillo-Lliva, S. & Malliri, A. The diverse roles of Rac signaling in tumorigenesis. *Cell Cycle* **10**, 1571–1581 (2011).
- Itoh, R. E. *et al.* Activation of Rac and Cdc42 video imaged by fluorescent resonance energy transfer-based single-molecule probes in the membrane of living cells. *Mol. Cell Biol.* **22**, 6582–6591 (2002).
- Waterman-Storer, C. M., Worthylake, R. A., Liu, B. P., Burridge, K. & Salmon, E. D. Microtubule growth activates Rac1 to promote lamellipodial protrusion in fibroblasts. *Nature Cell Biol.* **1**, 45–50 (1999).
- Lingle, W. L. *et al.* Centrosome amplification drives chromosomal instability in breast tumor development. *Proc. Natl Acad. Sci. USA* **99**, 1978–1983 (2002).
- Zhu, F. *et al.* The mammalian SPD-2 ortholog Cep192 regulates centrosome biogenesis. *Curr. Biol.* **18**, 136–141 (2008).
- Stehbens, S. & Wittmann, T. Targeting and transport: how microtubules control focal adhesion dynamics. *J. Cell Biol.* **198**, 481–489 (2012).
- Padua, D. & Massague, J. Roles of TGF $\beta$  in metastasis. *Cell Res.* **19**, 89–102 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We are grateful to S. Jhaveri-Schneider, S. Muthuswamy, M. Bettencourt-Dias, R. Basto, D. Calado, N. Ganem, A. Spektor, M. Kwon and B. Atkins for comments or discussion of the manuscript; the Nikon Imaging Center at Harvard Medical School; J. Waters, J. Rosenberg and H. Elliott from the Image and Data Analysis Core at Harvard Medical School for their help with FRET microscopy and analysis; L. Cameron of the Confocal and Light Microscopy Facility at Dana-Farber Cancer Institute; H. Li of SNP analysis; A. Bui for help establishing the 3D cultures; Q. Tseng for establishing MCF10A doublets on micropatterns; and C. Sproat and J. Marshall for help with the organotypic cultures. Reagents were kindly provided by L. Pelletier, J. Locarek, M. Matsuda, A. Hall and J. Marshall. S.A.G. is supported by an FCT grant HMSP-CT/SAU-ICT/0075/2009. M.T. is supported by the ERC grant 310472 and the ISI program of BPIFrance. D.P. is an HHMI investigator and is supported by an NIH grant GM083299-1.

**Author Contributions** S.A.G. and D.P. designed the experiments and wrote the manuscript. S.A.G. conceived, conducted and performed data analysis for most experiments. R.P. conceived and conducted all FRET experiments and Fig. 3a. M.T. and M.B. contributed with micro-pattern fabrication, Fig. 3d, Extended Data Fig. 6e, d and Extended Data Fig. 8c–e. R.D. contributed with Fig. 2b and Extended Data Fig. 5a. Y.S., K.P., C.T.L. and J.S.B. provided assistance with 3D cultures. All authors contributed with discussions and edited the manuscript.

**Author Information** The SNP data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through accession number GSE55042. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A.G. (s.godinho@qmul.ac.uk) or D.P. (david\_pellman@dfci.harvard.edu).

## METHODS

**Cell culture.** Human mammary epithelial MCF10A cells were maintained at 37 °C with 5% CO<sub>2</sub> atmosphere and cultured as previously described<sup>31</sup>. Briefly, MCF10A cells were grown in DMEM/F12 (Invitrogen) supplemented with 5% donor horse serum (Sigma), 20 ng ml<sup>-1</sup> epidermal growth factor (EGF; Sigma), 10 µg ml<sup>-1</sup> insulin (Invitrogen), 100 µg ml<sup>-1</sup> hydrocortisone (Sigma), 1 ng ml<sup>-1</sup> cholera toxin (Sigma), 100 U ml<sup>-1</sup> penicillin and streptomycin (Invitrogen). The MCF10A cell line overexpressing the human *ERBB2* gene (MCF10A.ErbB2 or MCF10A.NeuN) was previously characterized<sup>32,33</sup>. For 3D cultures, cells were grown in the same medium with reduced horse serum (2%) and EGF (5 ng ml<sup>-1</sup>). To assay invasion in 3D cultures, cells were grown in a mix of Matrigel:collagen-I, as previously described<sup>34</sup>. The addition of collagen-I to Matrigel facilitates invasion by increasing matrix stiffness<sup>35</sup>. We used growth factor-reduced Matrigel (BD Biosciences) lots with protein concentrations between 9 and 11 mg ml<sup>-1</sup>. Collagen-I (BD Biosciences) was used at 1.6 mg ml<sup>-1</sup>. Cells were grown for 4 days in 3D cultures before quantification of invasion. Between 200–300 acini were scored per condition for each experiment. To assess collagen-I degradation in 3D cultures, we added to the Matrigel:collagen-I mix 25 µg ml<sup>-1</sup> of quenched DQ-collagen-I (Molecular Probes). After degradation, DQ-collagen-I becomes fluorescent.

To collect cells from 3D cultures, we incubated cells with Dispase (BD Biosciences) for 30–60 min at 37 °C, according to the manufacturer's instructions. Cell aggregates were then trypsinized to obtain single cells suspensions and processed either for western blotting or to prepare chromosome spreads.

The alveolar epithelial cell line 16HBE, provided by Alan Hall, was cultured in MEM (Invitrogen) supplemented with GlutaMAX, Earle's salts, 10% of FBS and 100 U ml<sup>-1</sup> penicillin and streptomycin<sup>36</sup>. The non-transformed keratinocyte line (HaCaT), primary breast fibroblasts (1492N) and skin fibroblasts (HDFs) were provided by John Marshall. HaCaTs and HDFs were cultured in DMEM supplemented with 10% of FBS and 100 U ml<sup>-1</sup> penicillin and streptomycin. 1492N were cultured in 50:50 Ham's F12:DMEM supplemented with 10% of FBS and 100 U ml<sup>-1</sup> penicillin and streptomycin. We used tetracyclin-free FBS (Hyclone) to grow the 16HBE and HaCaTs cells expressing the *PLK4* construct to inhibit *PLK4* expression in the absence of doxycycline.

**Organotypic culture system.** The organotypic culture system was adapted from previously described methods<sup>37,38</sup>. Briefly, 200 µl of culture media containing epithelial cells or epithelial cells plus fibroblasts was added on top of the polymerized Matrigel:collagen-I mixture in each well. Co-culture of epithelial cells with fibroblasts was carried out as follows: HaCaTs cells were grown with HDF fibroblasts and MCF10A cells were grown with human primary fibroblasts (1492N). 600 µl of culture media was added to the bottom of the transwell. Fibroblasts invade the Matrigel:collagen-I layer whereas most of the non-invasive epithelial cells grow on top of this layer. Medium was changed every 2 days. Cells were grown for 7 days before samples were fixed and stained.

Cells in transwells were fixed in formalin overnight at room temperature. Inserts were removed using scalpels and placed in 70% ethanol and processed for histology. Paraffin embedded inserts were sectioned and stained for haematoxylin and eosin (H&E). For both MCF10A and HaCaTs, the percentage invasion was calculated as the number of cells that entered the Matrigel:collagen-I layer relative to the total number of cells per each well.

**Lentiviral and retroviral vectors.** For the generation of the inducible *PLK4* over-expression system we used the lentiviral vectors pLenti-CMV-TetR-Blast (17492, Addgene) and pLenti-CMV/TO-Neo-Dest (17292, Addgene)<sup>39</sup>. Wild-type *PLK4* and *PLK4*<sup>1-608</sup> cDNAs were cloned using the Gateway system into the pLenti-CMV/TO-Neo-Dest vector. Cells were first infected with a lentivirus containing the TetR and selected with Blasticidin (10–5 µg ml<sup>-1</sup>). After selection, cells were infected with the lentivirus containing the wild-type *PLK4* or *PLK4*<sup>1-608</sup> transgenes and selected with Geneticin (200–100 µg ml<sup>-1</sup>). Note that no clones were selected at any point and that all the selected cells were pooled to make a population. All the cell lines generated were induced with 2 µg ml<sup>-1</sup> of doxycycline for 48 h to induce the expression of the transgenes. pLKO.1 lentiviral vector containing Cep192 shRNA hairpin sequence was obtained from the RNAi consortium (TCR) at the Broad Institute (5'-CCCGG GAGGCATCAGTTAATACTGAT-CTCGAG-AT CAGTATTAAGTATGCTCTCTTTT-3'). pLKO.1 lentiviral vector expressing p53 shRNA was obtained from Addgene (19119)<sup>40</sup>. Lentilox Centrin1-eGFP construct was a gift from J. Loncarek. pLenti6/V5 lentiviral vector expressing H2B-GFP was cloned as previously described<sup>2</sup>. The Raichu-Rac1 biosensor, a gift from M. Matsuda, was subcloned into the retroviral vector pWZL-blast using the restriction sites EcoRI and SalI.

**Chemicals.** Doxycycline (Sigma) was used at 2 µg ml<sup>-1</sup>. The following doses of inhibitors were used: 25 µM NSC23766 (EMD Millipore), 50 µM CK-666 (Sigma), 0.1 µM Reversine (Cayman Chemical), 10 mM paclitaxel (Sigma), 4 µM of dihydrocytochalasin B (DCB; Sigma), 200 ng ml<sup>-1</sup> doxorubicin (Sigma), and 5 µM and 10 µM of marimastat (BB-2516; Sigma).

**2D indirect immunofluorescence microscopy.** Cells plated in glass coverslips were washed in PBS and fixed with 4% of PFA for 15 min at room temperature (RT). For centriole/centrosome staining, cells were fixed with ice-cold methanol at -20 °C for 10 min. Following fixation, cells were permeabilized with PBS and 0.2% Triton X-100 for 5 min, blocked in PBST (PBS, 5%BSA, 0.1% Triton X-100) for 30 min, and then incubated with primary antibodies in PBST for 60 min. Cells were washed with PBS and incubated with species-specific fluorescent secondary antibodies (Alexa-conjugated, Molecular Probes). DNA was stained with Hoechst 33342 (1:5,000; Invitrogen) for 5 min in PBS. Coverslips were mounted with ProLong Antifade mounting medium (Molecular Probes). Antibodies used included anti  $\alpha$ -tubulin DM1a (1:500; Sigma-Aldrich), anti-centrin2 (1:100; Santa Cruz), anti  $\gamma$ -tubulin GTU88 (1:500; Sigma-Aldrich) and anti  $\beta$ -catenin (1:500; Abcam). Phalloidin was used to stain F-actin (1:250; AlexaFluor 568; Invitrogen). Images were collected with a Yokogawa CSU-22 spinning disk confocal mounted on a Zeiss Axiovert microscope using 404, 488 and 561 nm laser light. Captured images from each experiment were analysed with Slidebook software (Intelligent Imaging Innovations). Measurement of centrosome number was performed in mitotic cells, ~100 were scored in each experiment.

**3D indirect immunofluorescence microscopy.** Immunofluorescence of 3D cultures was carried out as previously described<sup>41</sup>. Briefly, cells grown in 3D cultures were washed with 1× PBS and fixed in 5% of formalin (Sigma) in PBS for 20 min at 37 °C. After fixation cells were rinsed 3 times, 10 min each, with PBS:glycine (100 mM) and permeabilized with 0.5% Triton X-100 in PBS for 10 min. Cells were blocked with 10% of goat serum (Sigma) in IF buffer (130 mM NaCl, 7 mM Na<sub>2</sub>HPO<sub>4</sub>, 3.5 mM NaH<sub>2</sub>PO<sub>4</sub>, 7.7 mM NaN<sub>3</sub>, 0.1% BSA, 0.2% Triton X-100, 0.05% Tween-20) for 1 h at room temperature and primary antibodies were incubated in the same solution overnight at 4 °C. Cells were washed 3 times, 20 min each, with IF buffer. When required, cells were incubated with secondary antibodies for 1 h at room temperature (Alexa-conjugated, Molecular Probes). Cells were washed twice with IF buffer and once with PBS followed by incubation with Hoechst 33342 (1:2,500; Invitrogen) for 20 min. 3D cultures were then mounted in ProLong Antifade mounting medium (Molecular Probes). Antibodies used included anti  $\alpha$ -tubulin FITC conjugated DM1a (1:100; Sigma), anti Laminin-V AlexaFluor 488 conjugated (1:100; Millipore), anti-acetylated-tubulin 611B-1 (1:100; Sigma), anti-fibronectin (1:100; BD Biosciences) and anti-Pericentrin (1:100; Abcam). Phalloidin was used to stain F-actin (1:100; AlexaFluor 568; Invitrogen). Images were collected with a Nikon A1R laser scanning confocal head mounted on a Nikon Ti-E motorized inverted microscope using 404, 488 and 561 nm laser light. Captured images from each experiment were analysed using NIS-Elements software (Nikon).

**Long-term live-cell imaging.** H2B-GFP expressing cells were grown on glass-bottom 12-well tissue culture dishes (MatTek) for 2D imaging or in Lab-Tek chambered coverglass for 3D imaging. Cells were imaged on a Nikon TE2000-E2 inverted microscope equipped with a cooled CCD camera (TE2000, Orca ER, Hamamatsu; Ti-E, Coolsnap HQ2, Photometrics), a precision motorized stage (Bioprecision, Ludl), and Nikon Perfect Focus. Microscope was enclosed within temperature- and CO<sub>2</sub>-controlled environments that maintained an atmosphere of 37 °C and 3–5% humidified CO<sub>2</sub>. GFP and bright field images were captured at multiple points every 4 min for 1–2 days with either 10× (0.3 NA) or 20× (0.5, 0.75 NA) objectives. Captured images from each experiment were analysed using NIS-Elements software.

**Immunofluorescence quantification.** Quantification of  $\gamma$ -tubulin fluorescence intensity associated with the centrosomes was performed as previously described<sup>41</sup>. Images were acquired with a 100× NA 1.45 Plan Apo objective. Stacks of 6 images with 0.6 µm step size were collected. Step size was calculated to have minimal pixel overlapping between steps. SUM intensity projections of the images were used to quantify fluorescence intensity using ImageJ. Computer generated 50 × 50 and 80 × 80 pixel regions were centred over each centrosomes (as shown in Extended Data Fig. 9). The intensity value measured for the 50 × 50 pixel region include both centrosome and background fluorescence. Background fluorescence was obtained by subtracting the integrated value of 50 × 50 pixel region from the larger 80 × 80 pixel region. Integrated centrosomal fluorescence intensity was calculated by subtraction of the background fluorescence intensity from total fluorescence intensity (Extended Data Fig. 9). The advantage of this approach is that it controls for the non-homogeneity in background fluorescence.

**2D cell-cell adhesion analysis.** MCF-10A cells expressing the Raichu-Rac (here used just to visualize the cell membrane), with and without centrosome amplification, were plated on glass-bottom tissue culture dishes (MatTek) coated with 20 µg ml<sup>-1</sup> fibronectin (Sigma-Aldrich) for 20 min. Centrosome amplification was induced by 48 h *PLK4* induction, as described above. To monitor the cell-cell contact dynamics, we acquired time-lapse series of images, visualizing the CFP moiety of the Raichu-Rac FRET reporter (Fig. 3a and Supplementary Videos 6 and 7) at 5 min intervals over 12.4 h. To illustrate differences in cell-cell contact dynamics, we generated kymographs (Fig. 3a). The kymograph was obtained by



sequentially mounting a  $1 \times 336$  (pixel) region of interest from the videos, using ImageJ. Time-lapse series were collected with a Nikon inverted microscope with epi-fluorescence optics using a  $40\times$  plan Apo NA 1.4 objective. The microscope was equipped with a Nikon Perfect Focus System and a Hamamatsu ORCA ER cooled CCD camera and controlled with Nikon NIS-Element software. To reduce illumination intensity, minimizing phototoxicity and photobleaching, we used a ND8 (1/8 transmission) neutral density filter.

**Western blotting.** Cells were collected and resuspended in Laemmli buffer and proteins were separated on sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto PVDF membranes. Antibodies used included anti  $\alpha$ -tubulin DM1a (1:2,000; Sigma), anti-CEP192 (1:1,000; gift from L. Pelletier), anti-MCAK (1:1,000; Bethyl Laboratories), anti-p53 (1:1,000, Cell Signaling), anti-P-p53 (1:1,000, phospho-Ser15, Cell Signaling), E-cadherin (1:1,000, Invitrogen) and anti-Rac1 (1:1,000, BD Biosciences). Images were acquired using ImageQuant LAS4000 (GE Healthcare) and when required band intensity was quantified using ImageJ.

**siRNA.** siRNA was performed using Lipofectamine RNAiMax (Invitrogen) according to the manufacturer's instructions. 50 nM of CEP192, MCAK and luciferase (negative control) siRNA was used per well in a 6-well plate. After 6 h of incubation, transfected cells were washed and normal growth medium was added. Cells were analysed at 48 h after transfection. ON-Target plus SMART pools were used for MCAK and CEP192 siRNA (Dharmacon). Human KIF2C/MCAK siRNA (L-004955-00-0005): GGCAUAGCUCUGUGAAU (J-004955); CCAACGCA GUAUGGUUUA (J-004955-07); GCAAGCAACAGGUGCAAGU (J-004955-08); UGACUGAUCCUAUCGAAGA (J-004955-09). Human CEP192 siRNA (L-032250-01-0005): UGUGAAGAAUACGAGAU (J-032250-09); GCUCAGC GGUAUUUGGAA (J-032250-10); GUCUAGAACUCGAGAAUCA (J-032250-11); GGUUGAAGCAGUAGAGAGU (J-032250-12).

**Single-nucleotide polymorphisms (SNP).** Cells from 3D cultures were recovered after 4 days as described above (see Cell Culture section for details) and genomic DNA was prepared using DNA purification Kit (Qiagen) according to the manufacturer's instructions. Genomic DNA was used to perform analysis with genome-wide human SNP 6.0 arrays to determine copy number in MCF10A control cells, MCF10A after induction of centrosome amplification and MCF10A depleted of MCAK. SNP array data was analysed with the Nexus Copy Number Software (BioDiscovery; <http://www.biodiscovery.com/software/nexus-copy-number/>), using the 'Matched Paired Analysis' module. The SNP data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus<sup>42</sup> and are accessible through GEO accession number GSE55042.

**Microtubule polymerization assay.** Cells plated in glass coverslips were transferred to ice cold medium and incubated 1 h on ice to depolymerize microtubules. Cells were then incubated with medium at  $37^\circ\text{C}$  for 30 s to allow microtubules to polymerize and fixed immediately in ice-cold methanol for 10 min. Cells were stained for microtubules and centrioles (as described above) and microtubule number was quantified manually. Images of cells used in the analysis were acquired from 2 independent experiments.

**qRT-PCR.** PLK4 is highly unstable and we were not able to monitor its transient overexpression by western blotting. qRT-PCR was therefore used to analyse the extent of its expression. RNA was prepared using Qiagen RNeasy kit according to the manufacturer's instructions. For 3D cultures, RNA was initially harvested using TRIzol, and after chloroform extraction, the upper aqueous phase was used to purify RNA using the RNeasy kit. 300 ng of RNA was used to produce cDNA using qScript cDNA SuperMix kit (Quanta Bioscience), according to the manufacturer's instructions. For qRT-PCR, we used Power SYBR Green followed by analysis with ViiA PCR machine (Applied Biosystems).

Primers used to assess the levels of PLK4 overexpression only amplify the exogenous wild-type PLK4 or PLK4<sup>1-608</sup> sequences. The primers used for qRT-PCR were: PLK4 forward: 5'-CAGGATTGCGCGGATGGCG-3'; PLK4 reverse: 5'-AACCACTGTGAATGGACTCAGCTCT-3'; GAPDH forward: 5'-TTAAAG CAGCCCTGGTGAC-3'; GAPDH reverse: 5'-CTCTGCTCCTCTGTTGAC-3'.

**Rac1-GTP pull-down.** The Rac1-GTP pull-down assay was performed using the Rac1 activation kit (Cytoskeleton) according to the manufacturer's instructions. We used cells plated in one 10 cm dish per assay. Cells were resuspended in 400  $\mu\text{l}$  of lysis buffer and 15  $\mu\text{l}$  of CRIB/PBD beads were used to pull-down active Rac1. Extracts were incubated with the beads for 30 min. All the procedures were done at  $4^\circ\text{C}$  and buffers were kept ice-cold. After washing, beads were resuspended in 15  $\mu\text{l}$  of Laemmli buffer and processed for western blotting. To inhibit microtubule dynamics, cells were treated with 10 mM of paclitaxel for 1 h<sup>26</sup>.

**Generation of tetraploid cells with normal centrosome number.** MCF10A cells were treated with 4  $\mu\text{M}$  DCD for  $\sim 18$  h, washed every 5 min over 30 min, and then FACS sorted by DNA content using Hoechst at 1:2,500 (Molecular Probes) to isolate tetraploid cells with extra centrosomes. To isolate tetraploid cells with normal centrosome number, cells with a DNA content of 8c (dividing tetraploid

cells) were isolated and cultured for  $\sim 8$  days before a second FACS sorting to re-isolate 8c cells. By sort 4, nearly 100% of tetraploid cells (as assessed by FACS and karyotyping) had two centrosomes<sup>2</sup>.

**Chromosome spreads.** MCF10A cells were treated with 20 ng ml<sup>-1</sup> colcemid (Gibco) for 4 h, trypsinized, resuspended in 75 mM of KCl and incubated for 30 min at  $37^\circ\text{C}$ . Cells were then fixed with 3:1 ice-cold methanol:acetic acid, (Carnoy's solution) pelleted, and then washed three times more with methanol:acetic acid before being dropped on a pre-cleaned glass slide. Cells were allowed to dry on the slide and were then stained for 3 min with Giemsa stain in  $1\times$  Gurr's buffer (Gibco). Following a wash in Gurr's buffer, coverslips were added to slides and sealed with Permount (Fisher). Images of spreads were taken with a  $100\times$  objective on a Zeiss upright microscope and chromosomes were counted manually using Adobe Photoshop. The chi squared test of independence for nominal variables was used to calculate the *P* value.

**FACS.** Cells were fixed with 70% ethanol at  $4^\circ\text{C}$  followed by incubation with 250  $\mu\text{g ml}^{-1}$  RNaseA and 10  $\mu\text{g ml}^{-1}$  propidium iodide (Invitrogen) at  $37^\circ\text{C}$  for 30 min. FACS analysis was performed with a FACSCalibur flow cytometer (Becton Dickinson) and data analysed with CellQuest software.

**Micropatterning.** Glass coverslip micropatterning was performed as previously described<sup>43</sup>. Coverslips were first spin-coated with adhesion promoter Ti Prime (MicroChemicals) and then with 1% polystyrene in toluene at 3,000 r.p.m. Polystyrene coated coverslips were oxidized through oxygen plasma (FEMTO; Diener Electronics) for 15 s at 30 W before incubating with 0.1 mg ml<sup>-1</sup> PLL-PEG in 10 mM HEPES pH 7.4 for 30 min. After drying, coverslips were exposed to deep ultraviolet (UV) (UVO cleaner, Jelight) through a photomask (TOPPAN) for 5 min. Right after UV activation, coverslips were incubated with 20  $\mu\text{g ml}^{-1}$  of fibronectin (Sigma), and 20  $\mu\text{g ml}^{-1}$  of fluorescent fibrinogen conjugate (Invitrogen) solution in PBS for 30 min. Coverslips were washed 3 times with sterile PBS before plating cells. Approximately 0.5 million cells were seeded onto micropatterned chips and were washed after 30 min to remove non-attached cells, which almost always resulted in single cells per micropattern. Cells were allowed to divide for 16 h prior to fixation. For Rac1 and Arp2/3 inhibition, cells were treated with NSC27633 (25  $\mu\text{M}$ ) or CK-666 (50  $\mu\text{M}$ ), respectively, for 6 h, before fixation. For CEP192 siRNA, cells were plated after 48 h of siRNA treatment. Images of cells used in the analysis were acquired from 2 independent experiments. The hourglass micropattern guides intracellular junction positioning as shown in Figure 3b, where most of the junction extremities are positioned in the region deprived of extracellular matrix (fibronectin). We measure the angle and size of the junctions to assess junction position and length in the different conditions.

Note that when the cells divide on these patterns, spindle orientation is completely random with respect to the geometry of the pattern. The cells also move after division before establishing normal cell-cell contacts, further altering the position of the prior division site relative to the site of eventual cell-cell contact. Thus, the final adhesion pattern is completely random relative to the initial (and random) orientation of the spindle, as previously described<sup>20</sup>.

**Rac1 FRET.** For FRET experiments we used MCF10A cells stably expressing Raichu-Rac<sup>25</sup>. Centrosome amplification was induced with Dox for 48 h and  $\sim 10,000$  cells per ml of control and extra-centrosomes cells were plated onto 20  $\mu\text{g ml}^{-1}$  fibronectin (Sigma-Aldrich) coated glass-bottom dishes/plates (MatTek) for 6 h. For EGF depletion experiments cells incubated without EGF for 15 h.

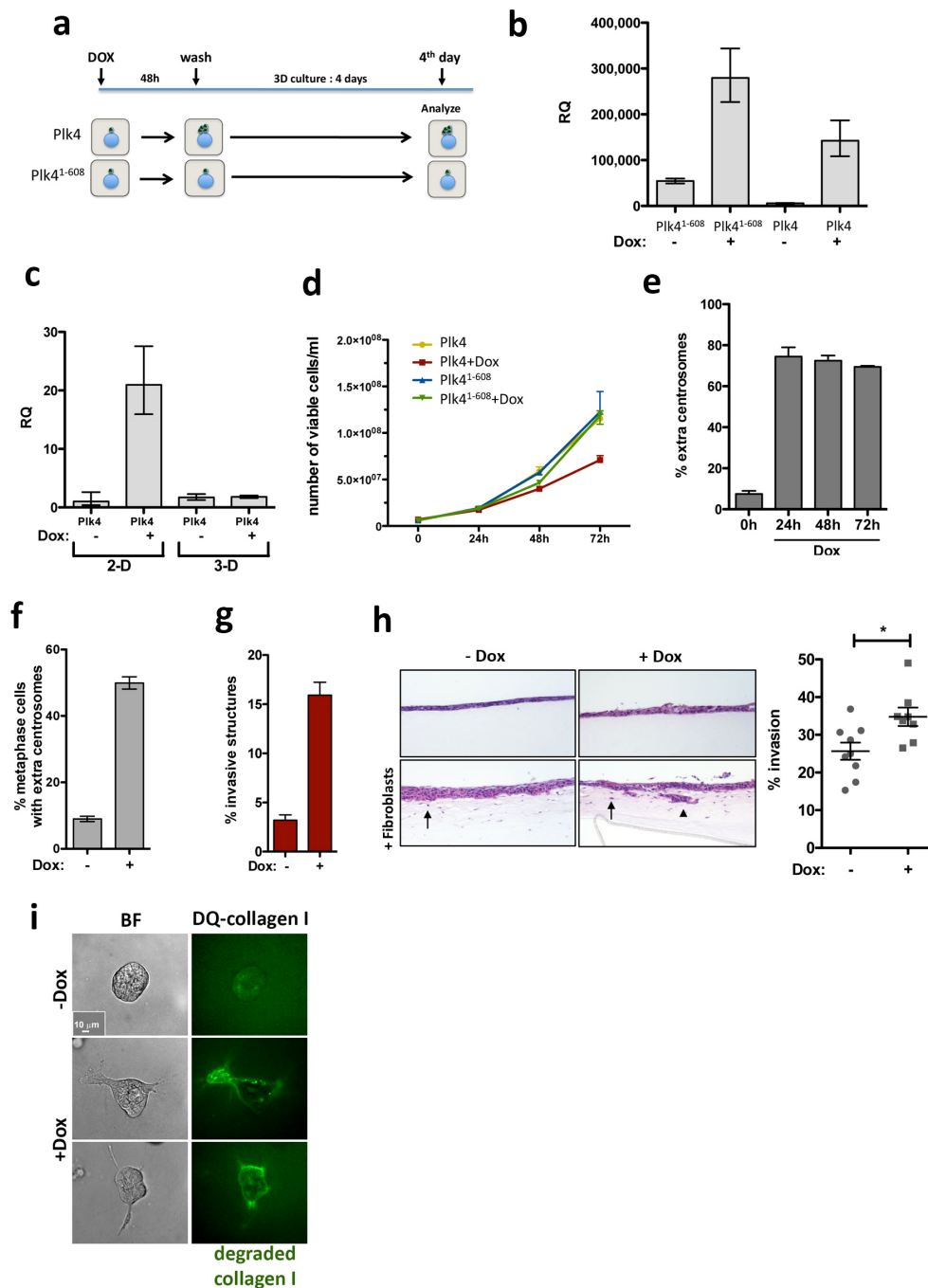
**FRET acquisition.** Before acquisition, to increase the signal to noise, the culture medium was exchanged with 199 medium (Life Technologies) without serum and phenol red. For FRET imaging, cells were excited using a Nikon Intensilight source with a 430/24nm (for CFP) excitation filter with a dual band pass (ECFP/EYFP #89002) dichroic mirror and 470/24 (for CFP) and 535/40 (for FRET and YFP) emission filters from Chroma (Bellows Falls, VT). To reduce illumination intensity and thus minimize phototoxicity and photobleaching, we used a ND8 (1/8 transmission) neutral density filter. All images were collected with Nikon Ti inverted microscope with epi-fluorescence optics equipped with a  $40\times$  plan Apo NA 1.4 object lens, perfect focus system and a Hamamatsu ORCA ER cooled CCD camera controlled with Nikon NIS-Element software.

The following steps were taken to calculate the average magnitude and spatial variation of the camera's noise for image correction of the acquired channels (CFP, FRET): (1) a sequence of 10 dark-current images were taken with 600 ms exposure time and  $4 \times 4$  binning as the experimental images, but with no light incident on the CCD; (2) a sequence of 10 shade/illumination correction images for each channel (CFP and FRET) were taken with 600 ms exposure time and  $4 \times 4$  binning as the experimental images, but acquired on a blank area without any objects. The average dark-current and shade images were calculated by averaging the acquired 10 images by using ImageJ. Single channel images (CFP, FRET) for each field view visualizing maximum four non-contacting cells were collected by using an exposure time of 600 ms,  $4 \times 4$  binning, a ND8 (1/8 transmission) neutral density filter and illumination light shuttered between acquisitions.

**FRET analysis.** The Biosensor software from Dr. Gaudenz Danuser's laboratory (<http://lccb.hms.harvard.edu/index.html>) along with a series of automated ImageJ macros were used to calculate the FRET ratio images. The acquired CFP and FRET images were corrected for the average dark-current, shading and background subtracted by using Biosensor or ImageJ. Then, single-cell images from each channel (CFP and FRET) were segmented by using the ImageJ minimum and mean threshold method, and the relative regions of interest (ROIs) were recorded. Single cell mask images having background and foreground (cell image) pixel values equal to zero and one, respectively, were calculated from each cell ROI by ImageJ. To set the image background pixels to zero, CFP and FRET images were multiplied by the corresponding calculated image masks. Finally the FRET ratio images were obtained by dividing the processed FRET image by the CFP image.

Single cell average FRET/CFP values representing the Rac1 activation levels were obtained by calculating the mean pixel value of the FRET ratio images for each ROI (single cell). FRET/CFP values were collected from 2 independent experiments. An unpaired two-tailed Student's *t*-test statistical analysis was used to obtain level of significance between different experiment conditions.

31. Debnath, J. *et al.* The role of apoptosis in creating and maintaining luminal space within normal and oncogene-expressing mammary acini. *Cell* **111**, 29–40 (2002).
32. Schafer, Z. T. *et al.* Antioxidant and oncogene rescue of metabolic defects caused by loss of matrix attachment. *Nature* **461**, 109–113 (2009).
33. Gunawardane, R. N. *et al.* Novel role for PDEF in epithelial cell migration and invasion. *Cancer Res.* **65**, 11572–11580 (2005).
34. Xiang, B. & Muthuswamy, S. K. Using three-dimensional acinar structures for molecular and cell biological assays. *Methods Enzymol.* **406**, 692–701 (2006).
35. Paszek, M. J. & Weaver, V. M. The tension mounts: mechanics meets morphogenesis and malignancy. *J. Mammary Gland Biol. Neoplasia* **9**, 325–342 (2004).
36. Wallace S. W., Magalhaes A. & Hall A. The Rho target PRK2 regulates apical junction formation in human bronchial epithelial cells. *Mol Cell Biol.* **31**, 81–91 (2011).
37. Gaggioli, C. *et al.* Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells. *Nature Cell Biol.* **9**, 1392–1400 (2007).
38. Nyström, M. L. *et al.* Development of a quantitative method to analyse tumour cell invasion in organotypic culture. *J. Pathol.* **205**, 468–475 (2005).
39. Campeau, E. *et al.* A versatile viral system for expression and depletion of proteins in mammalian cells. *PLoS ONE* **4**, e6529 (2009).
40. Godar, S. *et al.* Growth-inhibitory and tumor-suppressive functions of p53 depend on its repression of CD44 expression. *Cell* **134**, 62–73 (2008).
41. Hoffman, D. B., Pearson, C. G., Yen, T. J., Howell, B. J. & Salmon, E. D. Microtubule-dependent changes in assembly of microtubule motor proteins and mitotic spindle checkpoint proteins at Ptk1 kinetochores. *Mol. Biol. Cell* **12**, 1995–2009 (2001).
42. Edgar, R., Domrachev, M. & Lash, A. E. Gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
43. Azioune, A., Carpi, N., Tseng, Q., Thery, M. & Piel, M. Protein micropatterns: a direct printing protocol using deep UVs. *Methods Cell Biol.* **97**, 133–146 (2010).
44. Zhan, L., Xiang, B. & Muthuswamy, S. K. Controlled activation of ErbB1/ErbB2 heterodimers promote invasion of three-dimensional organized epithelia in an ErbB1-dependent manner: implications for progression of ErbB2-overexpressing tumors. *Cancer Res.* **66**, 5201–5208 (2006).
45. Yuan, K. *et al.* Primary cilia are decreased in breast cancer: analysis of a collection of human breast cancer cell lines and tissues. *J. Histochem. Cytochem.* **58**, 857–870 (2010).
46. Sonnen, K. F., Gabryjonczyk, A. M., Anselm, E., Stierhof, Y. D. & Nigg, E. A. Human Cep192 and Cep152 cooperate in Plk4 recruitment and centriole duplication. *J. Cell Sci.* **126**, 3223–3233 (2013).

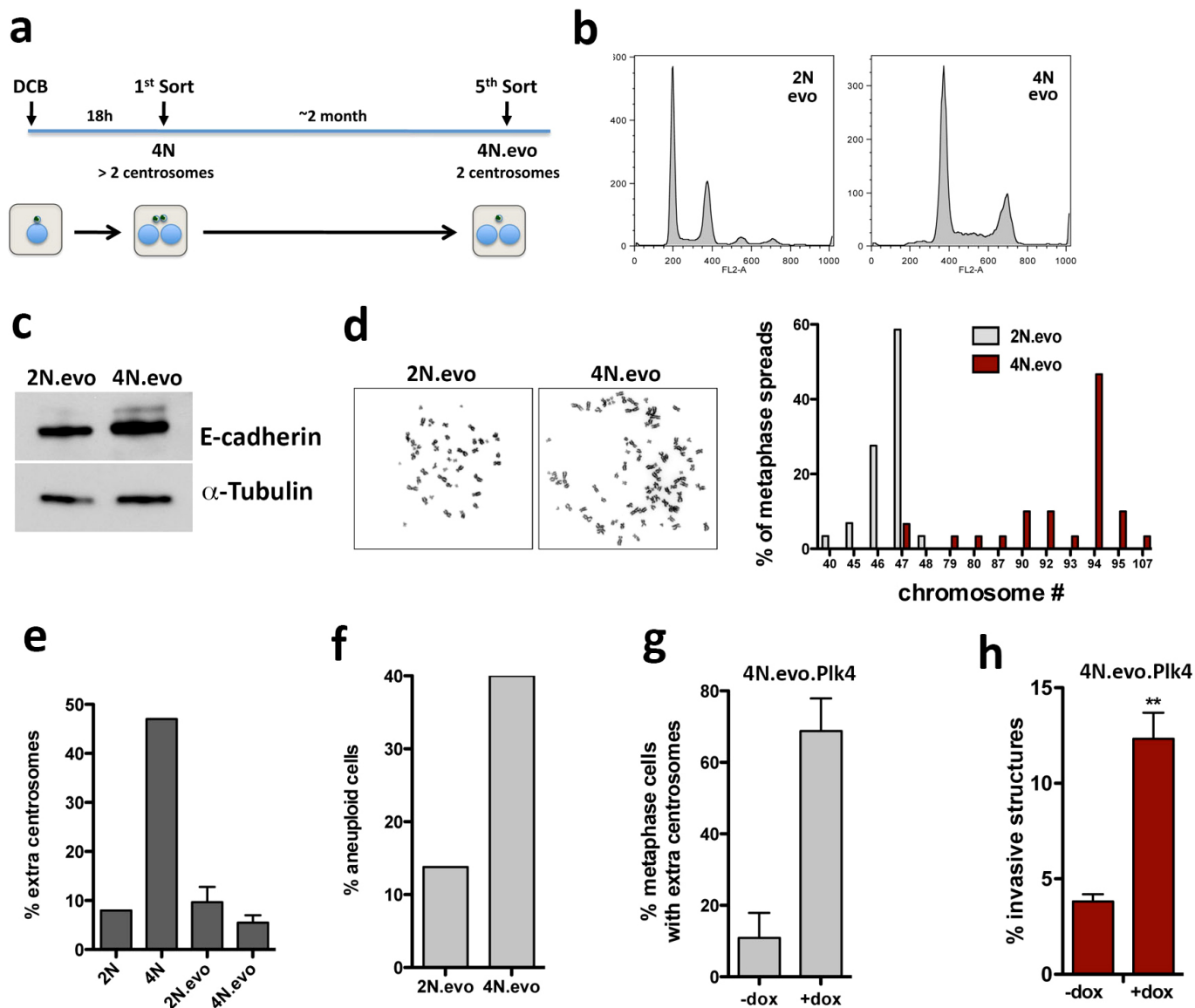


### Extended Data Figure 1 | Characterization of cells after transient overexpression of PLK4 or PLK4<sup>1-608</sup>

**a**, Scheme of the experimental design to induce centrosome amplification. Transient overexpression of *PLK4* and *PLK4*<sup>1-608</sup> was achieved by addition of Dox for 48 h (2D culture) followed by removal of Dox and growth in 3D culture for 4 days in the absence of Dox. **b**, qRT-PCR showing the levels of induction of the *PLK4*<sup>1-608</sup> and wild-type *PLK4* transgenes in cells after 48 h of Dox. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **c**, qRT-PCR showing the expression of *PLK4* after 48 h of Dox (2D) and after 4 days in 3D cultures (3D). Note that *PLK4* overexpression after 4 days in 3D cultures is down to control levels after Dox removal. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **d**, Proliferation curve of cells after induction of *PLK4* and *PLK4*<sup>1-608</sup> over 72 h. Centrosome amplification decreases cell proliferation. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **e**, The fraction of cells with centrosome amplification at the indicated time points after *PLK4* induction. Note that, because centrosome number is quantified in mitotic cells, this result demonstrates that cells with extra centrosomes can enter mitosis even after 72 h of Dox treatment. Error bars represent mean  $\pm$  s.e.

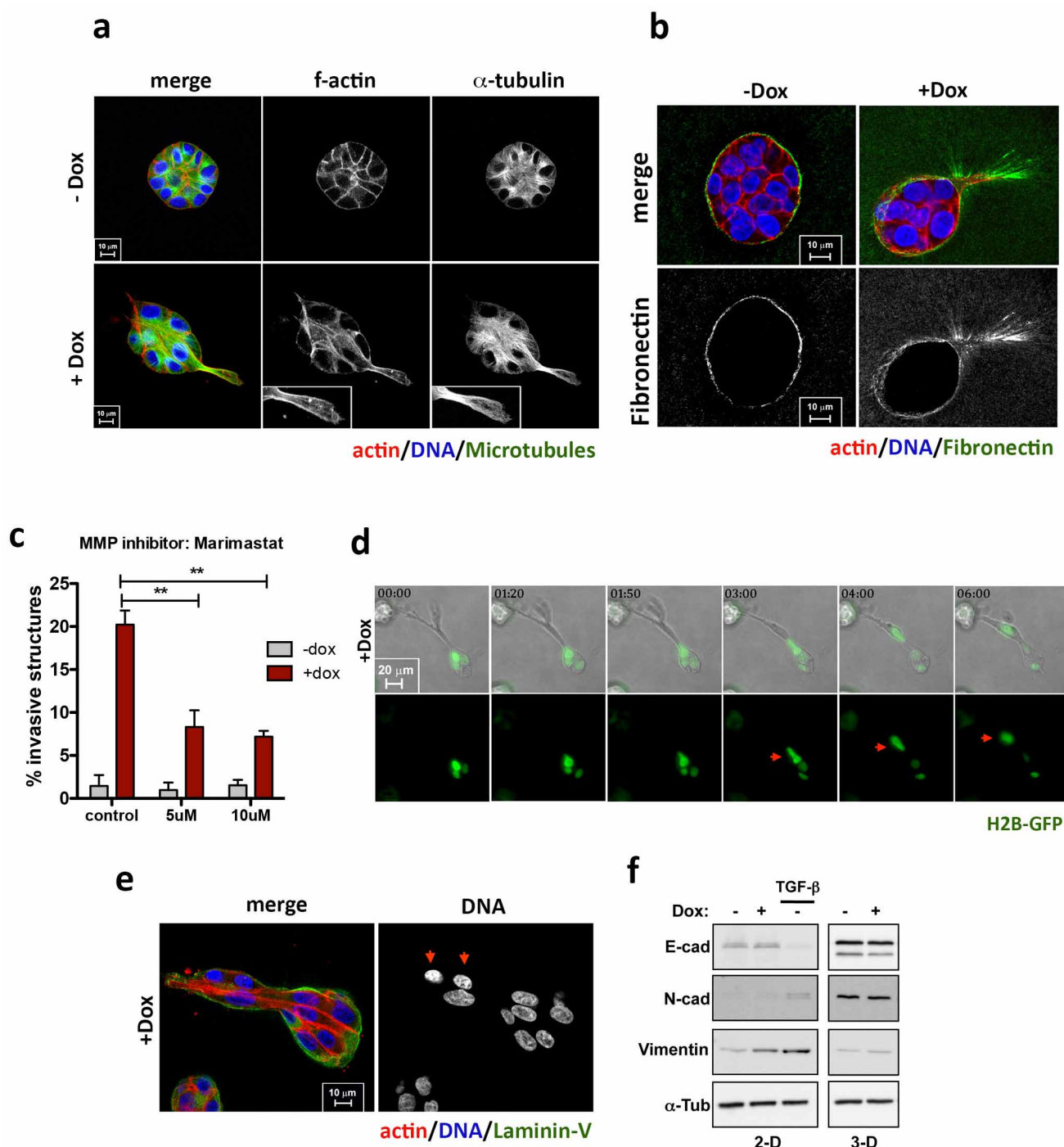
from 3 independent experiments. **f**, Fraction of cells with centrosome amplification in an independently generated MCF10A.*PLK4* cell line. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **g**, Corresponding fraction of invasive acini in 3D cultures. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **h**, Centrosome amplification (*PLK4* overexpression, +Dox) in non-transformed keratinocytes (HaCaTs) promotes invasion in the organotypic culture model. Images show H&E staining of sections of HaCaTs cells. Black arrows indicate cells invading the matrix. Note that the invasion of groups of cells was only detected in the +Dox condition (black arrowhead). Scale bar, 100  $\mu$ m. Graph shows quantification of the percentage of cells that invade. Each dot in the graphic represents the percentage of invasion in each individual well per experiment analysed. The *P* value was derived from unpaired two-tailed Student's *t*-test \**P* < 0.05. **i**, Collagen-I degradation induced by centrosome amplification (green). Collagen degradation is visualized by DQ-Col-I, which becomes fluorescent after degradation because of fluorophore dequenching. Scale bar, 10  $\mu$ m. See Methods for more details.





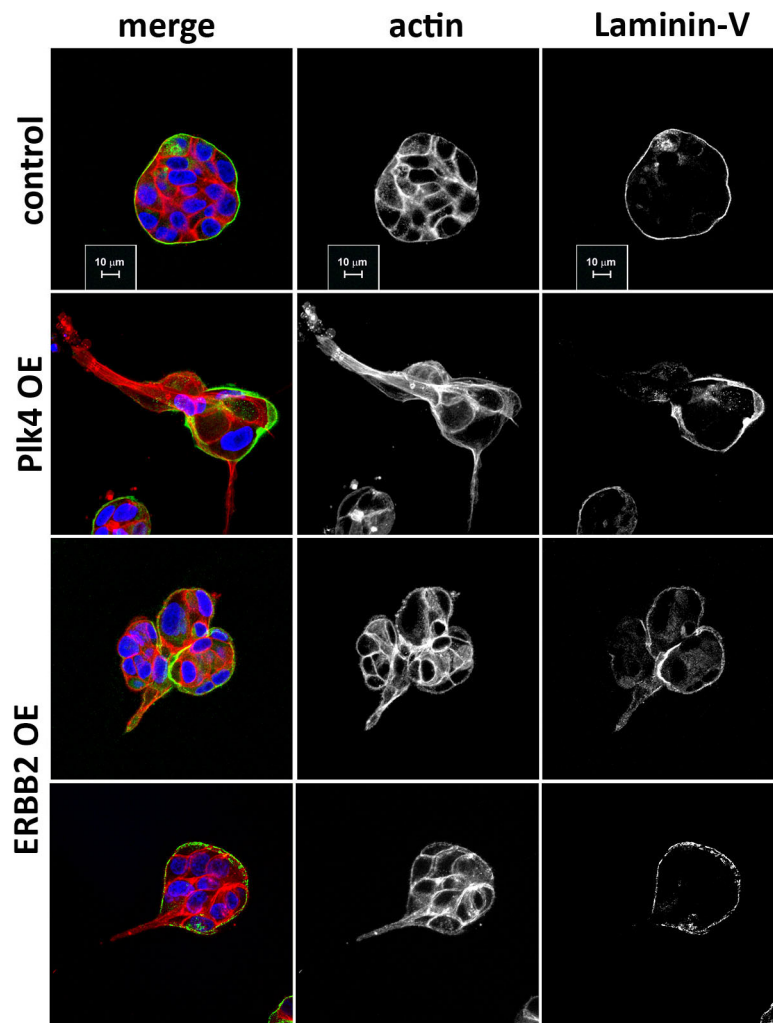
**Extended Data Figure 2 | Characterization of evolved diploid and tetraploid cells.** **a**, Scheme of the experimental design to obtain fresh MCF10A tetraploid cells with extra centrosomes (4N) and 'evolved' tetraploid cells that lost the extra centrosomes (4N.evo), as previously described<sup>2</sup>. **b**, FACS profiles of 'evolved' diploid (2N.evo) and tetraploid cells (4N.evo). **c**, Western blotting to detect E-cadherin in the 'evolved' cells indicates that 4N.evo maintain epithelial characteristics. **d**, Representative images of metaphase chromosome spreads of 2N.evo and 4N.evo and quantification of chromosome number by karyotyping (~30 chromosome spreads were quantified in each condition). 4N.evo cells have a near-tetraploid karyotype. **e**, Centrosome amplification in diploid cells (2N or 2N.evo) newly generated tetraploid cells (4N) and evolved tetraploid cells (4N.evo). **f**, Quantification of the

percentage of aneuploid cells in the 'evolved' cells. The 4N.evo cells are aneuploidy despite their near-tetraploid genomes (~30 chromosome spreads were quantified in each condition). **g**, Quantification of centrosome amplification of 4N.evo cells overexpressing PLK4. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **h**, Quantification of the invasive acini in 4N.evo cells after PLK4 overexpression. This experiment serves as a control to demonstrate that the 4N.evo cells retain their ability to amplify centrosomes and, after centrosome amplification, retain the capacity to form invasive acini. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. *P* value derived from unpaired two-tailed Student's *t*-test (\*\**P* < 0.005).



**Extended Data Figure 3 | Characterization of invasive structures in cells with extra centrosomes.** **a**, F-actin and microtubules in invasive protrusions: F-actin (red), microtubules ( $\alpha$ -tubulin, green) and DNA (blue). Insets show higher magnification images of the invasive protrusions. Scale bar, 10  $\mu$ m. **b**, Fibronectin at invasive protrusions: cells were stained for F-actin (red), fibronectin (green) and DNA (blue). Scale bar, 10  $\mu$ m. **c**, Fraction of invasive acini in 3D cultures after treatment with the broad spectrum matrix metalloprotease (MMP) inhibitor, marimastat (BB-2516). Error bars represent mean  $\pm$  s.e. from 3 independent experiments. *P* value derived from unpaired two-tailed Student's *t*-test (\*\**P* < 0.005). **d**, Images from videos of PLK4 overexpression cells (Supplementary Video 3), showing nuclei (labelled with H2B-GFP) migrating into an invasive protrusion (red arrows). Time scale

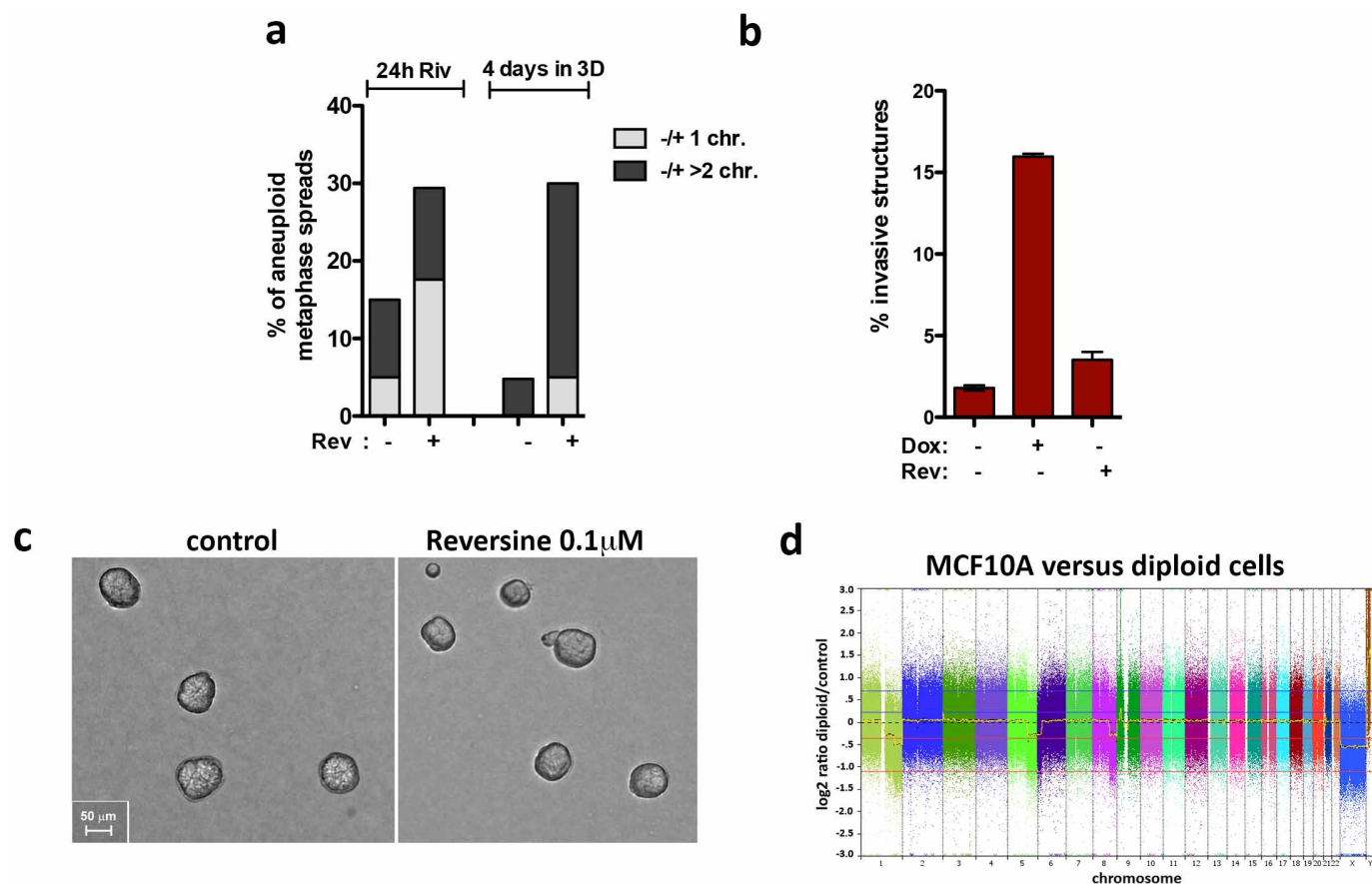
shown as h:min. Scale bar, 20  $\mu$ m. **e**, Multiple cells can migrate into invasive protrusions. Cells were stained for F-actin (red), laminin-V (green) and DNA (blue). Red arrows mark cells that migrated into the invasive protrusion. Scale bar, 10  $\mu$ m. **f**, Western blot showing levels of E-cadherin, N-cadherin and vimentin in cells with (+Dox) and without (–Dox) extra centrosomes before and after 4 days in 3D culture. The western blots show that, unlike cells treated with TGF- $\beta$ , cells with extra centrosomes do not acquire a canonical epithelial–mesenchymal transition (EMT) phenotype. We do note a small increase in the levels of vimentin in cells with extra centrosomes before plating in 3D cultures. Dox treatment was given for 48 h before 3D cultures in all experiments.



**Extended Data Figure 4 | Similarity between cells with centrosome amplification and cells with oncogene-induced invasion.** Cells were stained for F-actin (red), laminin-V (green) and DNA (blue). Similarity between the invasive protrusions of cells with extra centrosomes and the ones generated

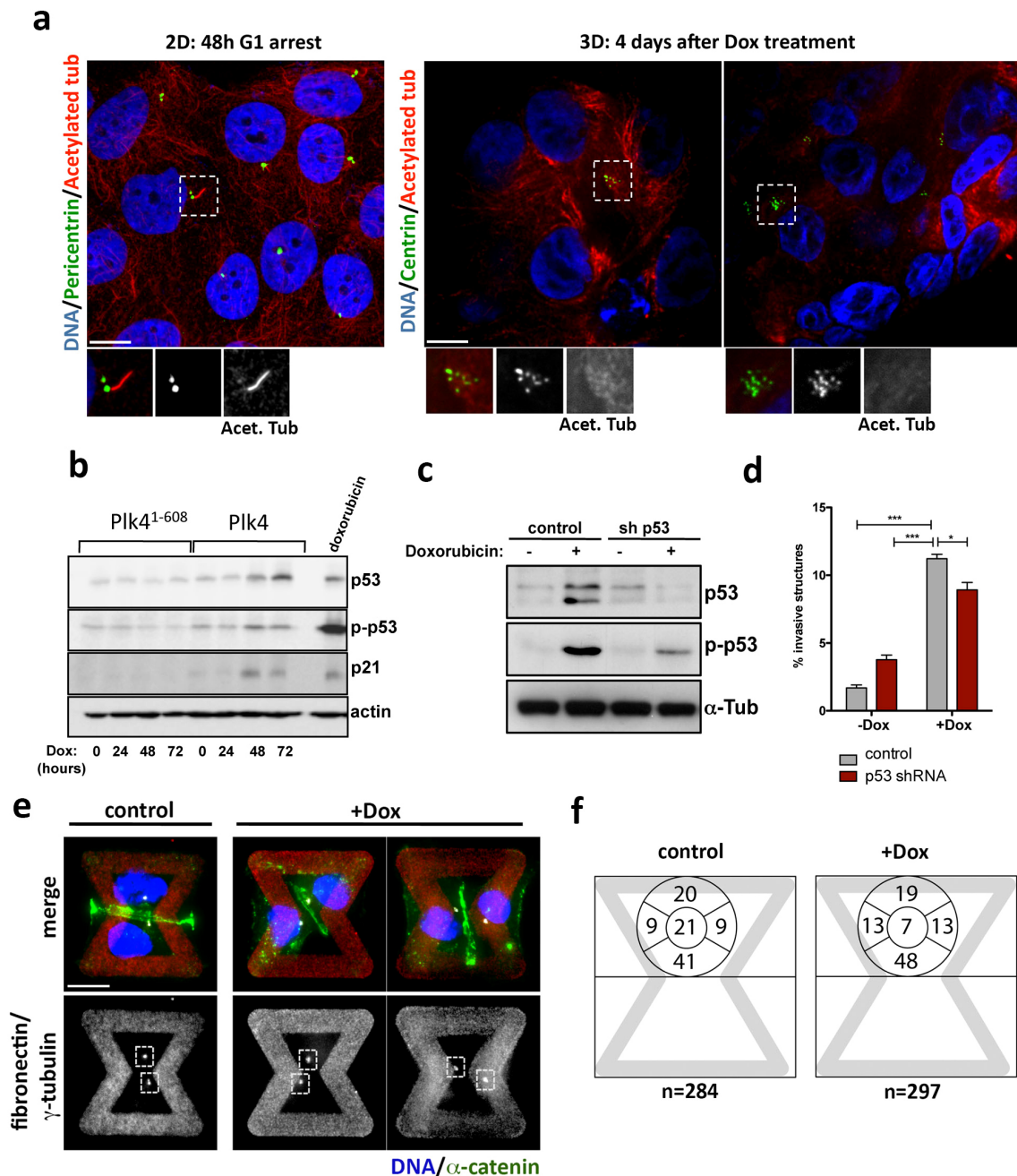
by cells overexpressing ERBB2, as previously reported<sup>4,44</sup>. In both conditions, invasive protrusions are characterized by the formation of actin-rich protrusions that are accompanied by degradation of the basement membrane. Scale bar, 10 μm.





**Extended Data Figure 5 | Reversine treatment induces aneuploidy but not invasive acini in MCF10A cells.** **a**, Quantification of the chromosome number in cells after treatment with with 0.1  $\mu$ M of reversine (Rev) for 24 h before and after 4 days in 3D cultures ( $\sim$ 20 chromosome spreads were quantified in each condition). The concentration of reversine used does not induce cytokinesis failure and therefore would not induce centrosome amplification by inducing tetraploidy. **b**, Fraction of invasive acini after MCF10A cells are treated with reversine. Increased aneuploidy from reversine

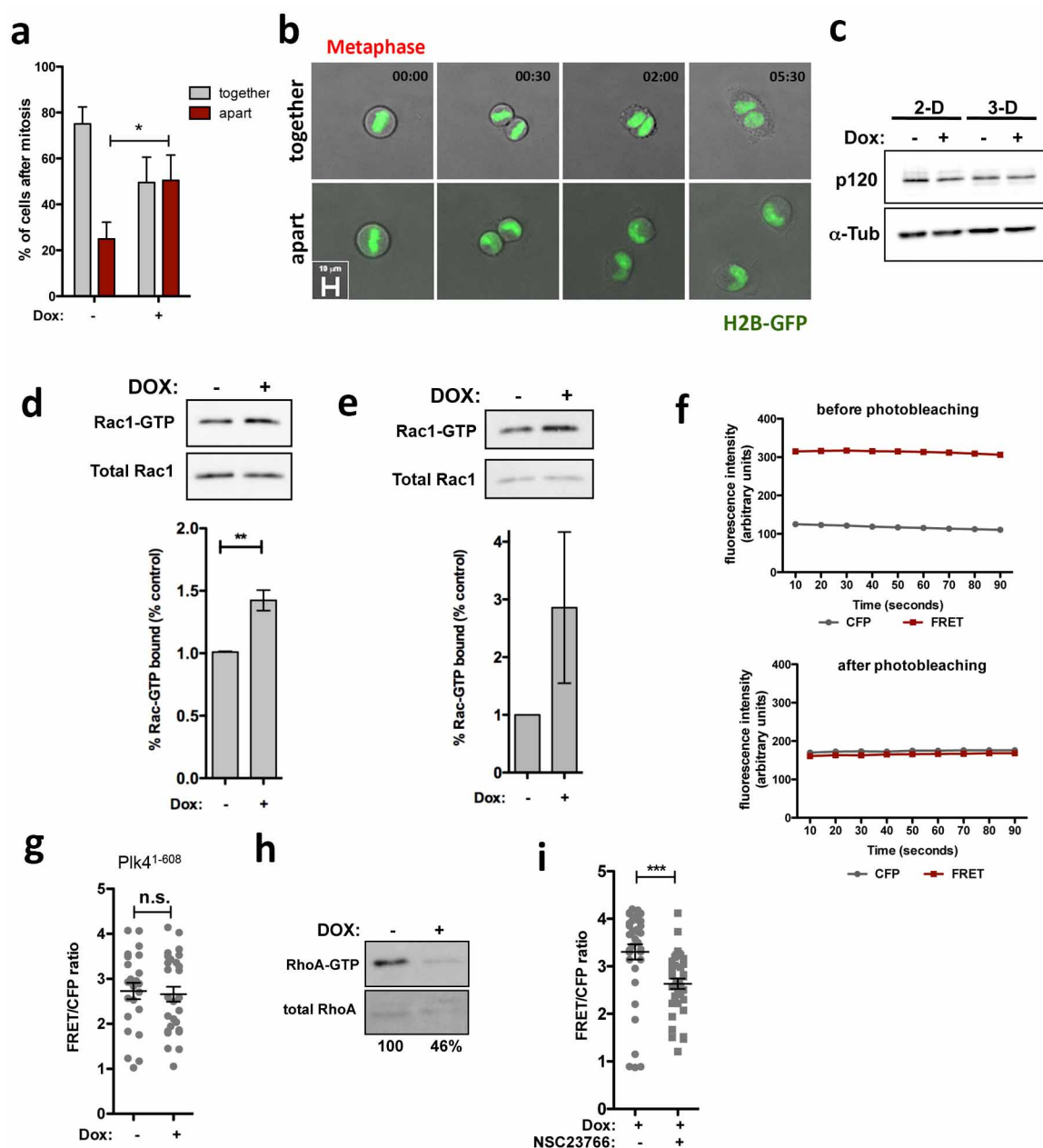
treatment does not induce invasion. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **c**, Images of 3D cultures after treatment with reversine showing normal appearing acini. Scale bar, 50  $\mu$ m. **d**, SNP analysis of MCF10A cells (without centrosome amplification) compared with Human Reference Genomic DNA 103 from Affymetrix. Previously reported genomic alterations in MCF10A cells can be detected in our analysis, namely: +5q, +6p and +8q.



**Extended Data Figure 6 | Invasive protrusions from 3D cultures of MCF10A cells with extra centrosomes are not an indirect consequence of altered cilia signalling, increased p53 expression or defects in centrosome polarization.**

**a**, Cells in 2D were stained for pericentrin (green, inset), acetylated tubulin (red, inset) and DNA (blue). Cells were arrested for 48 h in G1 to induce primary cilium formation. Note that even in this case most of the cells do not form cilia. This is expected because MCF10A cells have limited proficiency for cilia formation, with only ~7% of the cells assembling cilia even after 7 days of serum starvation<sup>45</sup>. Cells in 3D were stained for centrin (GFP, green inset), acetylated tubulin (red, inset) and DNA (blue). Cells do not form cilia after 4 days in 3D cultures. This is expected because, unlike MDCK cells, MCF10A cells do not have a discernable apical polarity and lumen after 4 days in 3D cultures and thus are unlikely to form primary cilium at this time. **b**, As expected<sup>17</sup>, centrosome amplification in MCF10A cells induces modest p53 activation. Note that this degree of p53 activation has a minor effect on the proliferation of MCF10A cells (Extended Data Figure 1d). Expression of PLK4 and PLK4<sup>1-608</sup> was induced by Dox for the indicated times: 0, 24, 48 and 72 h. **c**, Western blotting showing the levels of induction of p53 after doxorubicin treatment (200 ng ml<sup>-1</sup> for 4 h) in control and p53-depleted cells, demonstrating that the p53 shRNA efficiently prevents p53 activation.

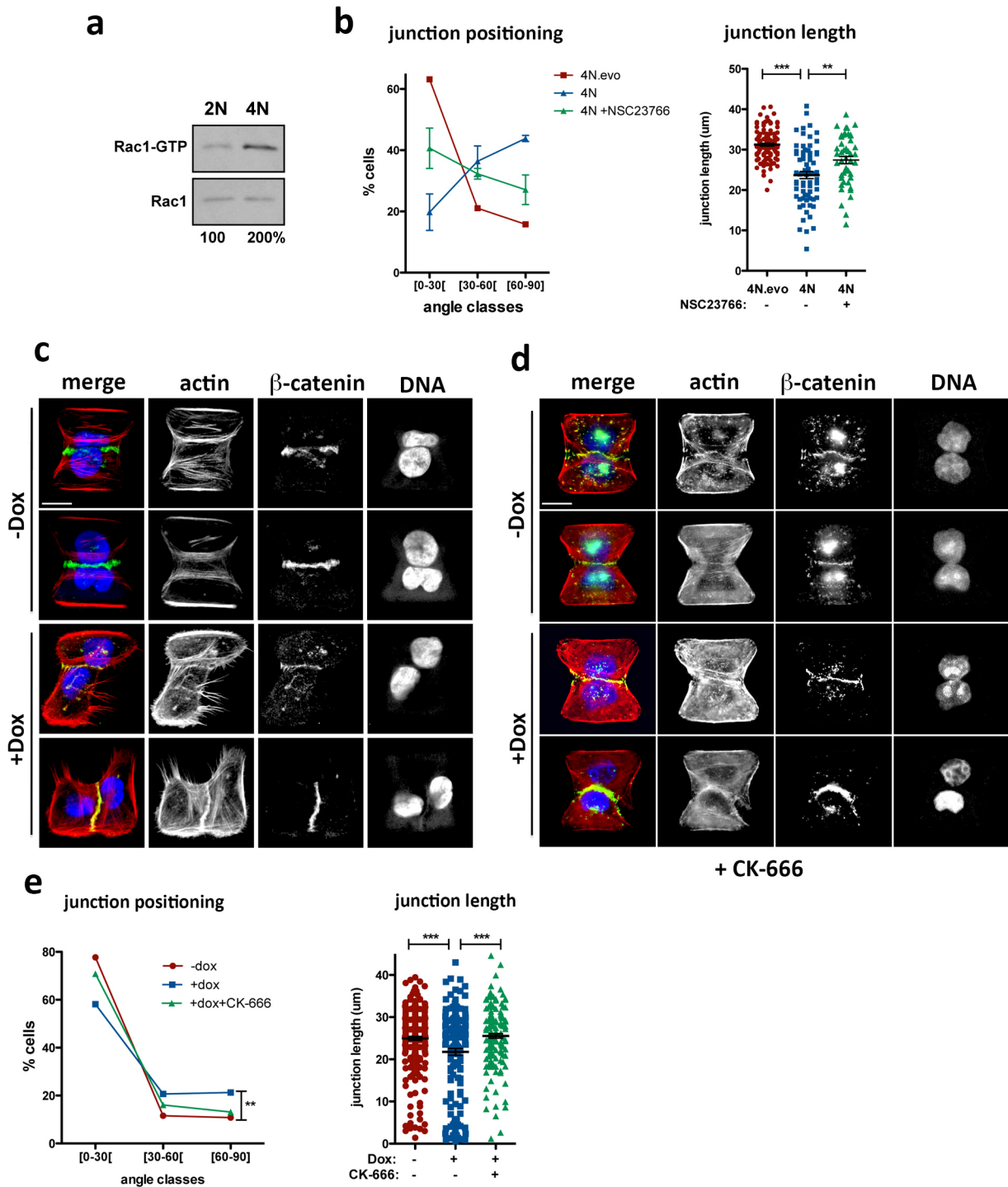
**d**, Fraction of acini with invasive protrusions from cells with (+Dox) or without (–Dox) centrosome amplification after depletion of p53. The ability of cells to form invasive acini is not significantly affected by their p53 status. Error bars represent mean ± s.e. from 3 independent experiments. *P* values were derived from unpaired two-tailed Student's *t*-test (\*\**P* < 0.0005; \**P* < 0.05). **e**, Cells were stained for α-catenin (green), γ-tubulin (red), DNA (blue) and the fibronectin micro-pattern visualized in red. Dashed boxes outline the centrosomes. Note that after PLK4 overexpression, extra centrosomes (clustered in interphase) are correctly positioned towards the cell–cell junction (similar to the control) even when the junction is defective, suggesting that centrosome amplification is not impairing the polarity axis of these cells. **f**, Centrosome positioning is not altered in cells with (+Dox) and without (control) extra centrosomes. Left image shows representative images of cells showing centrosomes (γ-tubulin) in relation to adherens junctions (α-catenin). Right image shows scheme with quantification of the fraction of centrosomes at the indicated positions on the micropatterns (see Fig. 3a). Note that the position of centrosomes in cells with centrosome amplification does not differ from that in control cells. Cells were plated on the patterns 48 h after induction of centrosome amplification with PLK4.



**Extended Data Figure 7 | Centrosome amplification induces cell scattering and Rac activation.** **a**, Quantification of number of cells with (+Dox) or without (−Dox) extra centrosomes that remain as pairs within 10 h time after mitosis. For −Dox  $n = 180$ ; +Dox  $n = 98$ . Cell scattering occurred in most of the cases within the first 2 h after mitosis. Cells were imaged on 2D substratum. Similar results were obtained with fixed cells. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **b**, Still images from videos showing examples of cells that stay together (−Dox) or move apart (+Dox). Time scale shown as h:min. **c**, Western blot showing levels of p120 catenin in cells with (+Dox) and without (−Dox) extra centrosomes in 2D and 3D cultures. **d**, Western blot from a pull-down experiment to detect GTP-bound Rac1 in HaCaT cells. Graph shows quantification of active Rac1 from pull-down experiments. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. **e**, Western blot from a pull-down experiment to detect GTP-bound Rac1 in 16HBE cells. Graph shows quantification of active Rac1

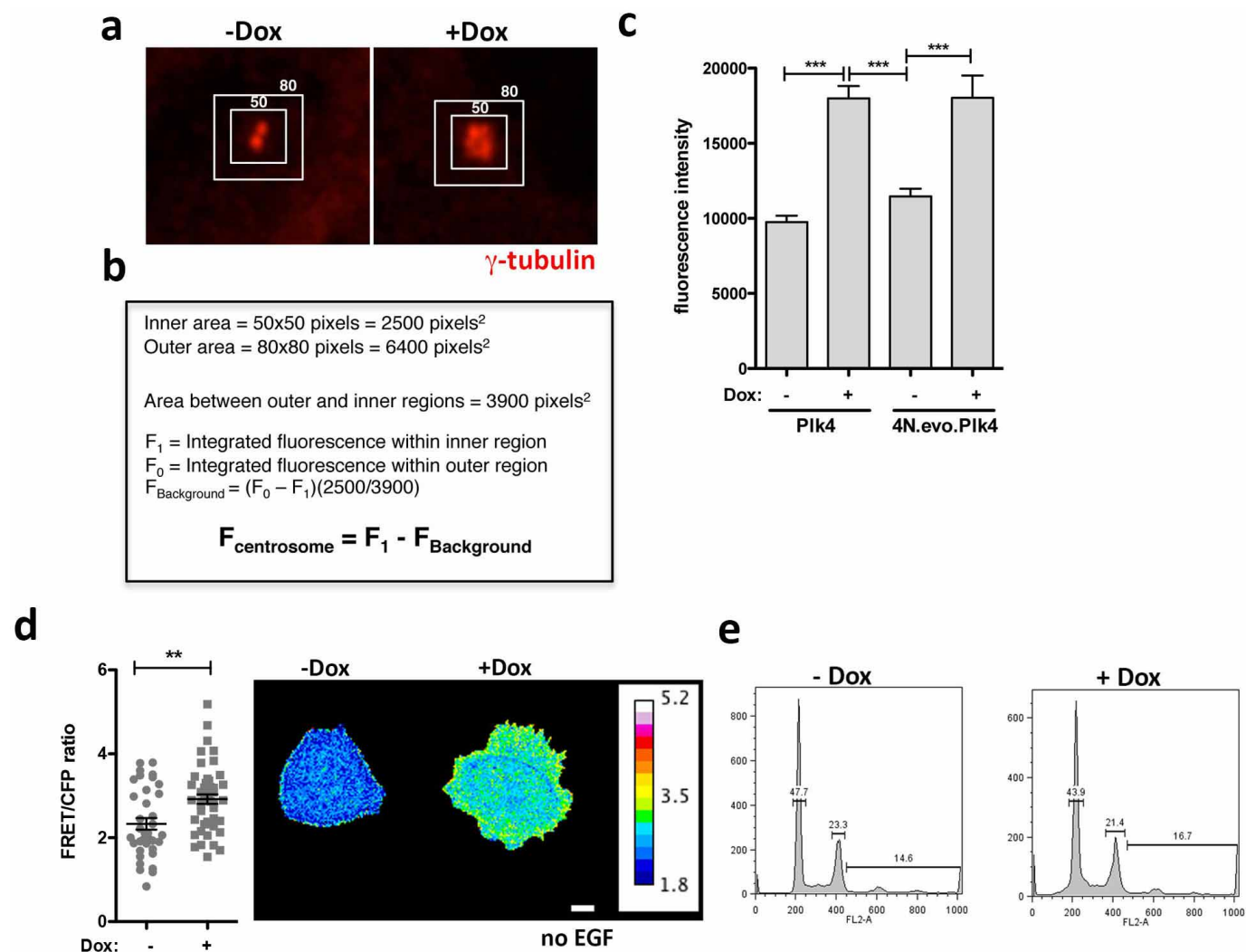
from pull-down experiments. Error bars represent mean  $\pm$  s.e. from 2 independent experiments. **f**, FRET control demonstrating increased CFP emission after photobleaching of the YFP fluorophore at an excitation wavelength of 510 nm for 10 min in MCF-10A single cells expressing Raichu-Rac. **g**, Levels of active Rac1 measured by FRET in cells overexpression PLK4<sup>1-608</sup>. For −Dox,  $n = 25$ ; +Dox,  $n = 22$ . Error bars represent mean  $\pm$  s.e. **h**, Western blot from a pull-down experiment to detect GTP-bound RhoA in MCF10A cells showing decrease RhoA activity in cells with extra centrosomes. **i**, Levels of active Rac1 measured by FRET in cells with extra centrosomes treated with the Rac1 inhibitor NSC23766, demonstrating that NSC23766 inhibits Rac1 activation in cells with extra centrosomes. For −NSC23766,  $n = 37$ ; +NSC23766,  $n = 36$ . Error bars represent mean  $\pm$  s.e. All the  $P$  values were derived from unpaired two-tailed Student's  $t$ -test (\*\* $P < 0.0005$ ; \* $P < 0.005$ ; \* $P < 0.05$ ). Scale bar, 10  $\mu$ m.





**Extended Data Figure 8 | Cell-cell adhesion defects caused by centrosome amplification can be observed in tetraploid cells and can be suppressed by Arp2/3 complex inhibition.** **a**, Western blot from a pull-down experiment to detect GTP-bound Rac1 in tetraploid MCF10A cells. **b**, Distribution of the cell-cell junction angles (left) and size (right) in the indicated tetraploid cells, with or without treatment with the Rac1 inhibitor, NSC23766. Note that tetraploid cells with extra centrosomes (4N) have a striking defect in junction positioning by comparison with tetraploid cells with normal centrosome number (4N.evo). This severe phenotype is only partially rescued by Rac1 inhibition. For 4N.evo,  $n = 106$ ; 4N,  $n = 70$ ; NSC23766,  $n = 47$ . Error bars represent mean  $\pm$  s.e. **c**, Examples of cell doublets with (+Dox) or without

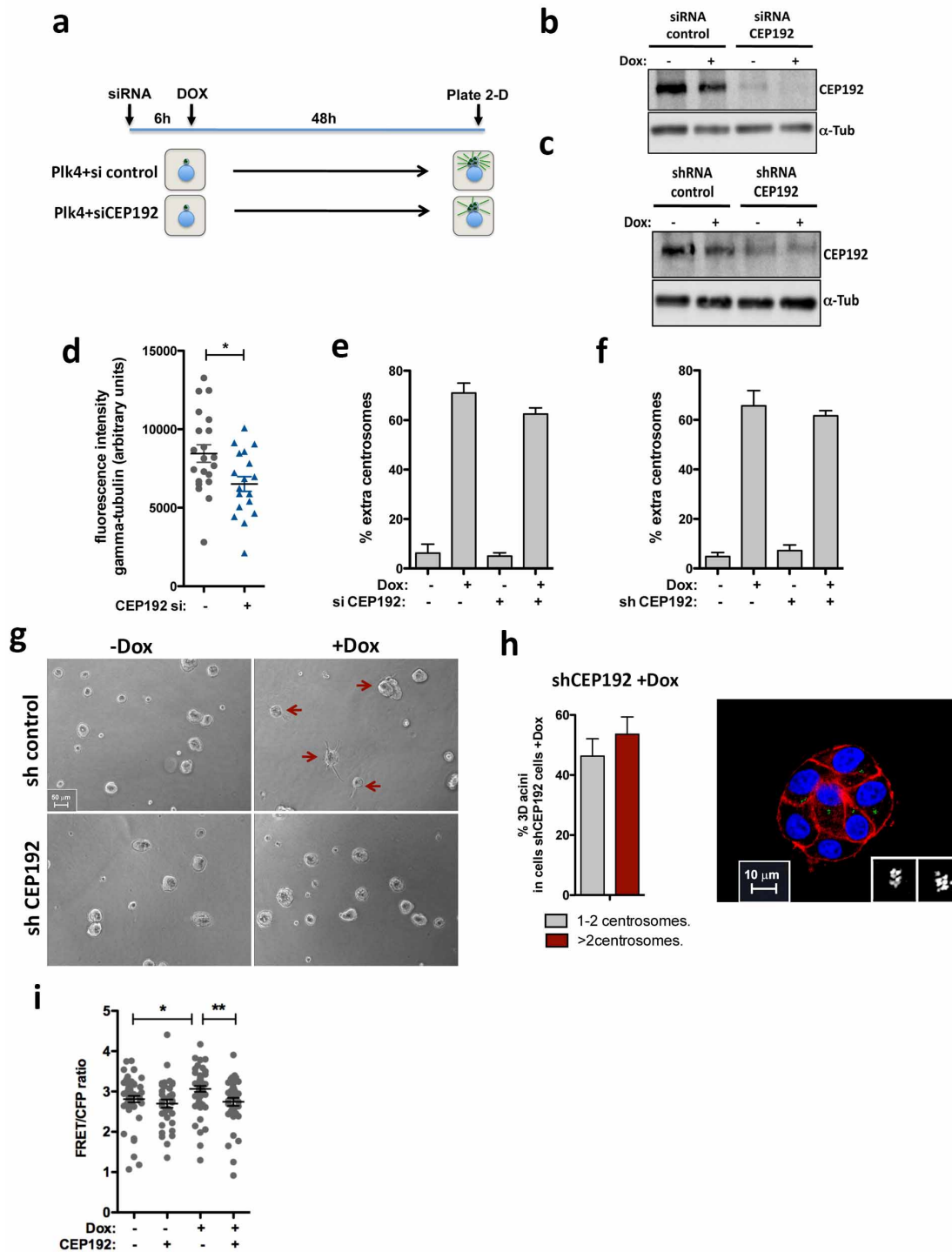
(-Dox) centrosome amplification on the fibronectin micro-patterns. Cells were stained for F-actin (red),  $\beta$ -catenin (green), DNA (blue). **d**, Examples of cell doublets with (+Dox) or without (-Dox) centrosome amplification on the fibronectin micro-patterns treated with the Arp2/3 inhibitor (CK-666). Cells were stained for F-actin (red),  $\beta$ -catenin (green), DNA (blue). **e**, Distribution of the junction angle and quantification of the junction size in cells with extra centrosomes treated with 50  $\mu$ M of Arp2/3 inhibitor (CK-666) for 6 h. Cells were analysed 48 h after Dox treatment. For -Dox,  $n = 251$ ; +Dox,  $n = 160$ ; CK666,  $n = 168$ . Error bars represent mean  $\pm$  s.e. All the  $P$  values were derived from unpaired Student's two-tailed  $t$ -test (\*\* $P < 0.0005$ ; \*\* $P = 0.005$ ). Scale bar, 10  $\mu$ m.



# Extended Data Figure 9 | Quantification of centrosomal $\gamma$ -tubulin in cells with extra centrosomes and levels of active Rac1 in the absence of EGF.

**a**, Images of centrosomes from interphase MCF10A cells stained for  $\gamma$ -tubulin. Boxes represent the region for measurement of centrosomal  $\gamma$ -tubulin signal (inside) and background (area between inner and outer boxes). **b**, Method used to determine the integrated fluorescent intensity of centrosomal  $\gamma$ -tubulin, as previously described<sup>41</sup>. **c**, Measurement of  $\gamma$ -tubulin intensity at the centrosomes showing that in interphase, increased centriole number is sufficient to increase  $\gamma$ -tubulin levels at the centrosomes whereas increased ploidy *per se* does not (4N.evo). For PLK4 -Dox,  $n = 60$ ; PLK4 +Dox,  $n = 49$ ; 4N.evo. PLK4 -Dox,  $n = 34$ ; 4N.evo. PLK4 +Dox,  $n = 35$ . Error bars represent mean  $\pm$  s.e. **d**, Increased Rac1 activity in cells with extra centrosomes can be

detected in arrested cells deprived of EGF. Quantification of Rac1 activity by FRET in single cells with (+Dox) and without (-Dox) extra centrosomes in the absence of EGF, and examples of FRET images of cells in the absence of EGF. For PLK4 -Dox,  $n = 36$ ; PLK4 +Dox,  $n = 47$ . Error bars represent mean  $\pm$  s.e. Scale bar, 10  $\mu$ m. **e**, FACS profiles of control (-Dox) and cells with extra centrosomes (+Dox) after 48 h of Dox treatment showing that there is not major difference in the cell cycle profiles of these cells. Note that at this time point centrosome amplification does not produce a striking defect in cell proliferation (Extended Data Fig. 1d). All the  $P$  values were derived from unpaired two-tailed Student's  $t$ -test (\*\* $P < 0.005$ ; \*\*\* $P < 0.0005$ ).



**Extended Data Figure 10 | Depletion of CEP192 suppresses the invasive properties of cells with centrosome amplification.** **a**, Scheme of the experimental design to induce centrosome amplification in cells depleted of CEP192 by siRNA. Transient overexpression of PLK4 is induced 6 h after siRNA to allow efficient centrosome overduplication. As expected, after depletion of CEP192 for 48 h, cells are partially compromised in their ability to overduplicate centrosomes after PLK4 overexpression<sup>46</sup>. **b**, Western blot showing efficient depletion of CEP192 after 48 h treatment of cells with CEP192 siRNA. **c**, Western blot showing partial depletion of CEP192 by shRNA. **d**, Quantification of centrosomal  $\gamma$ -tubulin after depletion of CEP192 by siRNA for 48 h. Similar results were observed with CEP192 esiRNA (not shown). It is notable that at least for a three-day period, cells remain viable after CEP192 knockdown. For control siRNA,  $n = 22$ ; CEP192 siRNA,  $n = 20$ . Error bars represent mean  $\pm$  s.e. Quantification of centrosome amplification after depletion of CEP192 by siRNA (**e**) or shRNA (**f**). Error bars represent

mean  $\pm$  s.e. from 3 independent experiments. **g**, Bright field images of acini after 4 days in 3D culture, demonstrating that partial CEP192 depletion by shRNA does not significantly impair cell growth or the formation of acini. Red arrows indicate the invasive acini. **h**, Quantification of PLK4-mediated centrosome amplification in cells depleted of CEP192 after 4 days in 3D cultures showing that these cells still carry extra centrosomes. Error bars represent mean  $\pm$  s.e. from 3 independent experiments. Images show normal acini displaying centrosome amplification after partial knockdown of CEP192. Cells were stained for F-actin (red), centrioles (centrin1-GFP, green) and DNA (blue). Scale bar, 10  $\mu$ m. **i**, Levels of active Rac1 measured by FRET after CEP192 depletion. For ctr.siRNA -Dox,  $n = 51$ ; ctr.siRNA +Dox,  $n = 35$ ; CEP.siRNA -Dox,  $n = 53$ ; CEP.siRNA +Dox,  $n = 37$ . Error bars represent mean  $\pm$  s.e. All the  $P$  values were derived from unpaired two-tailed Student's  $t$ -test (\* $P < 0.05$ ; \*\* $P < 0.005$ ).



# Membrane proteins bind lipids selectively to modulate their structure and function

Arthur Laganowsky<sup>1\*</sup>, Eamonn Reading<sup>1\*</sup>, Timothy M. Allison<sup>1</sup>, Martin B. Ulmschneider<sup>2</sup>, Matteo T. Degiacomi<sup>1</sup>, Andrew J. Baldwin<sup>1</sup> & Carol V. Robinson<sup>1</sup>

Previous studies have established that the folding, structure and function of membrane proteins are influenced by their lipid environments<sup>1–7</sup> and that lipids can bind to specific sites, for example, in potassium channels<sup>8</sup>. Fundamental questions remain however regarding the extent of membrane protein selectivity towards lipids. Here we report a mass spectrometry approach designed to determine the selectivity of lipid binding to membrane protein complexes. We investigate the mechanosensitive channel of large conductance (MscL) from *Mycobacterium tuberculosis* and aquaporin Z (AqpZ) and the ammonia channel (AmtB) from *Escherichia coli*, using ion mobility mass spectrometry (IM-MS), which reports gas-phase collision cross-sections. We demonstrate that folded conformations of membrane protein complexes can exist in the gas phase. By resolving lipid-bound states, we then rank bound lipids on the basis of their ability to resist gas phase unfolding and thereby stabilize membrane protein structure. Lipids bind non-selectively and with high avidity to MscL, all imparting comparable stability; however, the highest-ranking lipid is phosphatidylinositol phosphate, in line with its proposed functional role in mechanosensation<sup>9</sup>. AqpZ is also stabilized by many lipids, with cardiolipin imparting the most significant resistance to unfolding. Subsequently, through functional assays we show that cardiolipin modulates AqpZ function. Similar experiments identify AmtB as being highly selective for phosphatidylglycerol, prompting us to obtain an X-ray structure in this lipid membrane-like environment. The 2.3 Å resolution structure, when compared with others obtained without lipid bound, reveals distinct conformational changes that re-position AmtB residues to interact with the lipid bilayer. Our results demonstrate that resistance to unfolding correlates with specific lipid-binding events, enabling a distinction to be made between lipids that merely bind from those that modulate membrane protein structure and/or function. We anticipate that these findings will be important not only for defining the selectivity of membrane proteins towards lipids, but also for understanding the role of lipids in modulating protein function or drug binding.

Three membrane protein complexes were selected to give a range of topologies, oligomeric states and anticipated selectivity towards lipids: (1) the pentameric MscL from *M. tuberculosis* with two transmembrane helices per subunit and an intimate relationship with lipids<sup>10,11</sup>; (2) the tetrameric water efflux channel AqpZ from *E. coli* with six transmembrane helices for which associated lipid or detergent molecules have been revealed in crystal structures and in related homologues<sup>12,13</sup>; and (3) the trimeric AmtB from *E. coli*, with eleven transmembrane helices involved in the transport of ammonia, for which no lipid binding has been observed to date in crystal structures<sup>14</sup>.

Our first objective for studying membrane protein and lipid interactions by mass spectrometry was to maintain the folded state of these multimeric assemblies in the gas phase, devoid of detergent from the micelles required for their introduction<sup>15</sup>, such that stabilization by lipid binding could be deduced (Fig. 1a). This had not been possible previously as the conditions needed to disrupt the micelle perturbed the folded

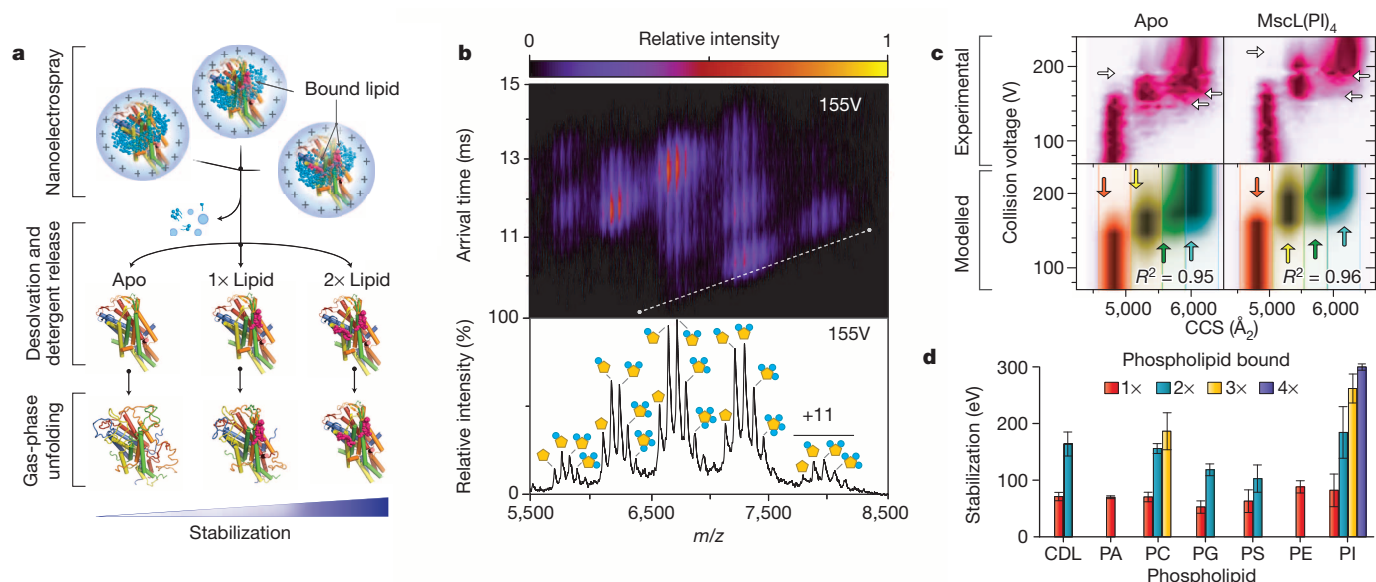
state of membrane protein complexes. We screened non-ionic detergents<sup>16</sup> and employed ion mobility mass spectrometry (IM-MS) (Supplementary Video 1), a technique that measures the rotationally averaged collision cross-sections (CCS) of protein complexes in the gas-phase<sup>17</sup>. We found that in detergents commonly used for structural studies (*n*-dodecyl- $\beta$ -D-maltoside (DDM), octyl glucose neopentyl glycol (OGNG) and *n*-nonyl- $\beta$ -D-glucopyranoside (NG)), the CCS of all charge states were substantially greater when emerged from the micelle than those calculated for the crystal structure, indicative of gas-phase unfolding (Extended Data Fig. 1). Interestingly, for tetraethylene glycol mono-octyl ether (C<sub>8</sub>E<sub>4</sub>) lower average charge states were observed with CCS in agreement with those calculated for the crystal structures, thus providing optimal conditions for mass spectrometry of intact membrane protein complexes.

To assess individual lipid-binding events, we prepared a series of synthetic and natural lipid solutions and added these to protein complexes in C<sub>8</sub>E<sub>4</sub> (Extended Data Fig. 2). MscL was tested first, revealing well-resolved mass spectral peaks corresponding to populations of MscL in complex with one to five lipids (Fig. 1b and Extended Data Fig. 3). Unlike other biophysical approaches, which report an average of all species in solution, whether bound or unbound, discrete lipid-bound states can be resolved and interrogated individually, allowing us to obtain CCS for individual lipid-binding events and to assess their effects on stability. Maintaining folded structures of these lipid-bound complexes enabled us to explore molecular dynamics simulations combined with filtering using CCS measurements<sup>18</sup> to locate probable lipid-binding sites (Extended Data Figs 4 and 5 and Supplementary Discussion). We then applied collision-induced unfolding, analogous to that used previously for studying ligand binding to soluble complexes<sup>19</sup>. An increase in CCS, from the trend line calculated for the folded state, is observed with increasing collision voltage and is assigned to highly extended conformers (Fig. 1b and Supplementary Video 2). All species, apo and one to five phospholipid-bound forms, are measured within the same experiment enabling a direct comparison, and show less unfolding of native pentameric states as a function of lipid binding.

To extract quantitative values for the effects of lipid binding, we applied an equilibrium unfolding model, used extensively in solution studies involving chemical denaturation<sup>20,21</sup>. Plotting CCS against collision voltage reveals unfolding trajectories from the native state, through intermediates, to extended forms. Importantly, the oligomeric state is maintained throughout the unfolding experiment (Extended Data Fig. 6 and Fig. 1c). We identified four distinct intermediate states in the unfolding trajectories and by comparing the apo form with lipid-bound states we calculated the stabilization imparted at the transitions by each lipid (see Methods). The stabilization computed for bound lipids allows us to compare and rank seven different lipids, as well as synthetic phosphatidylcholine (PC) lipids with chain lengths varying from C14 to C24. Despite differences in these lipid structures, we found that all stabilized MscL to a similar degree, consistent with a previous fluorescence study with engineered Trp residues<sup>22</sup> (Fig. 1d and Extended Data Fig. 3b,c). We also found that MscL avidly bound phosphatidylinositol phosphate

<sup>1</sup>Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 5QY, UK. <sup>2</sup>Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA.

\*These authors contributed equally to this work.



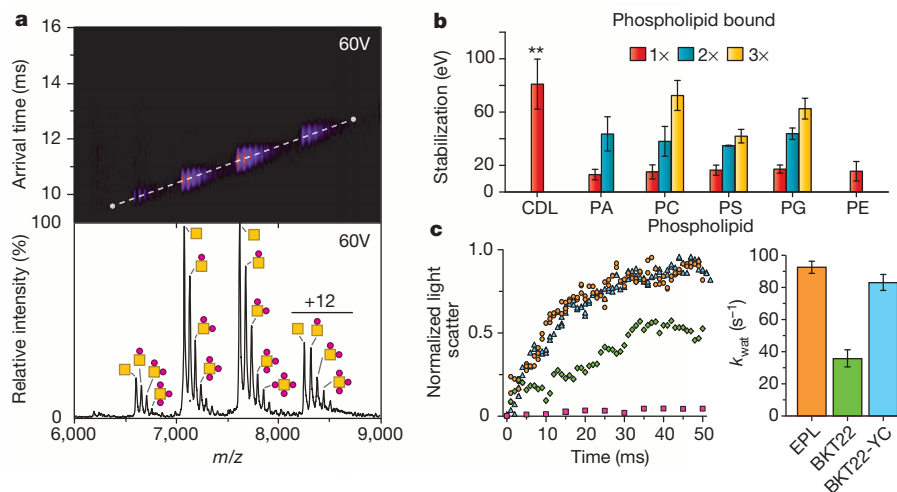
**Figure 1 | The mechanosensitive channel of large conductance (MscL) resists unfolding in the presence of lipids.** **a**, MscL in electrospray droplets, within a lipid-detergent micelle, undergoes desolvation and activation. **b**, Mass spectrum reveals several MscL phosphatidylinositol phosphate (PI) bound states. Ion mobility (top) of MscL(PI)<sub>0-5</sub> with a trend line for native state (dotted). **c**, Plot of collision voltage against CCS for +12 ions of apo and

MscL(PI)<sub>4</sub>. Experimental and modelled unfolding plots with collision voltages at which transitions occur and CCS values (horizontal and vertical arrows, respectively).  $R^2$  values are provided. **d**, Stabilization calculated from parameters defined by fitting MscL (+12) with lipids. Reported are average and s.e.m. from repeated measurements ( $n = 3$ ) in units of electron volts (eV).

(PI), more strongly than all other lipids investigated, conferring a large linear increase in stability upon binding many PI molecules.

We applied a similar procedure to AqpZ, revealing well-resolved lipid binding and a CCS in agreement with that of the crystal structure of AqpZ (Fig. 2a and Extended Data Fig. 7). Monitoring unfolding trajectories for lipid-bound states of AqpZ showed similar transition points between the different states that were largely independent of the bound lipid. This implies that the lipids investigated had similar effects on resistance to unfolding (Fig. 2b and Supplementary Video 3). As successive lipids are added, from one to three, stabilization is increased via regular increments consistent with a linear, cumulative effect. The only exception

to this is cardiolipin (CDL), for which binding resulted in a substantial increase in protein stability compared to all other lipids. Intrigued by this observation, we performed functional assays<sup>23</sup> of the channel reconstituted into liposomes of varying lipid composition. Water permeability assays of AqpZ reconstituted in a total polar lipid extract from a CDL-deficient *E. coli* strain<sup>24</sup> (BKT22) showed a striking reduction in water transport compared to the channel reconstituted in wild-type *E. coli* polar lipid preparations (EPL) (Fig. 2c and Extended Data Fig. 8). Moreover, reconstituting AqpZ in a lipid preparation from strain BKT22 expressing cardiolipin synthase genes<sup>24</sup>, restores CDL to near wild-type levels (mole percentage of 6.6 versus 9.1), and results in water transport



**Figure 2 | AqpZ is indiscriminately stabilized by lipids with the exception of cardiolipin, a lipid that stabilizes the channel significantly and modulates its function.** **a**, Mass spectra reveal AqpZ(POPC)<sub>0-5</sub> resolved at 60 V with IM-MS arrival times in agreement with the CCS calculated from the crystal structure. **b**, Stabilization of AqpZ (+13) bound to lipids, calculated from unfolding parameters. Shown as described in Fig. 1d. One-way ANOVA

( $n = 3$ ),  $**P < 0.01$ . **c**, Water permeability assay for AqpZ reconstituted in total polar lipid extracts from *E. coli* (EPL) (orange), or a cardiolipin-deficient strain<sup>24</sup> (BKT22) (green), or BKT22 cells expressing ClsC and YmdB to restore cardiolipin (BKT22-YC) (cyan), compared with empty EPL liposomes (pink) (Extended Data Fig. 8). Rate constants of water transport ( $k_{\text{wat}}$ ) and standard error of replicates ( $n = 5$ ).

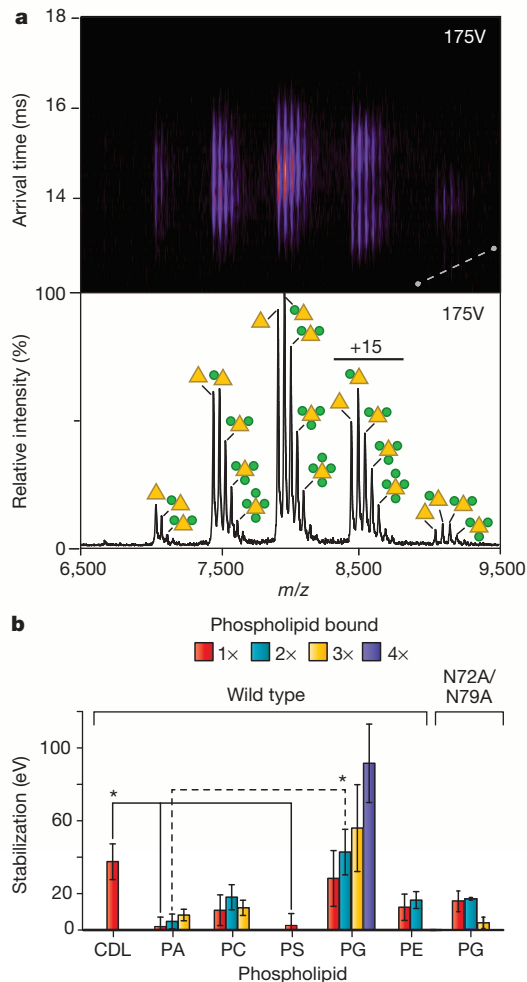
activity comparable to in EPL (Fig. 2c). These results clearly demonstrate CDL modulates the function of AqpZ.

Analogous IM-MS experiments carried out on AmtB show that resistance to gas-phase unfolding following binding of anionic lipids, phosphatidic acid (PA) and phosphatidylserine (PS), is not correlated with significant stabilization of the channel (Fig. 3). Zwitterionic lipids, PC and phosphatidylethanolamine (PE), confer only moderate stabilization. By contrast, addition of CDL or phosphatidylglycerol (PG) results in striking increases in stabilization with cumulative binding of PG increasing protein stability linearly (Fig. 3b, Extended Data Fig. 7 and Supplementary Video 4). These results therefore reveal a defined selectivity for PG-like headgroups.

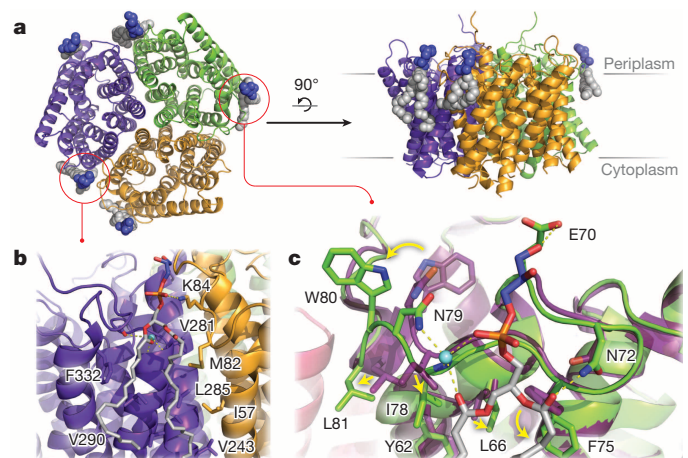
To understand the origin of this observed selectivity, we crystallized AmtB in the presence of PG. A screen produced X-ray-grade crystals which were used for structure determination at 2.3 Å resolution (Extended Data Table 1). The structure of AmtB is composed of three protein subunits, each comprising of 11 transmembrane helices<sup>14</sup>, with two trimers located in the asymmetric unit (Fig. 4). We resolved eight PG molecules in the electron density on the periplasmic side of AmtB (Extended Data

Fig. 9), a potential ninth PG molecule having ambiguous electron density. Two of the PG molecules are located in similar positions at the subunit interface, with headgroups not in contact with the protein (Fig. 4b). The overall structure is similar to the 20 previously reported in the Protein Data Bank (PDB) determined from detergent-only environments. Major differences arise from a distinct conformational change in a loop, residues 70–81, forming a specific lipid-binding site (Fig. 4c, Supplementary Video 5 and Supplementary Discussion). PG forms hydrogen bonds to N72 and E70 and a water bridge to N79. F75 flips downward to interact with a lipid tail of PG. In addition, several residues are repositioned to interact with the phospholipid bilayer. L81 and I78 shift into the hydrophobic environment of the lipid bilayer. Unexpectedly, W80 moves almost 4 Å ( $C\alpha$  to  $C\alpha$ ), its side-chain rotated 160° compared to other structures. This rotation relocates the W80 side-chain from a protein environment into one that is ideally positioned to interact with headgroups of phospholipids, consistent with their alignment along the lipid bilayer interface (Extended Data Fig. 9d)<sup>25</sup>. Furthermore, mutation of N72 and N79 to alanine (AmtB<sup>N72A/N79A</sup>), engineered to abolish the specific lipid-binding site, results in the loss of PG induced gas-phase stability (Fig. 3b), consistent with our IM-MS approach identifying specific lipid-binding events.

By resolving individual lipid-bound states of membrane protein complexes within the mass spectrometer, and interrogating their stability through gas-phase unfolding, we obtained quantitative values to rank lipid-binding interactions. MscL binds lipids non-selectively, without regard to particular headgroup or chain length, and binding of any lipid imparts comparable stability, in accord with the ability of the channel to bind and respond promiscuously to lipid composition<sup>11,26</sup>. The most significant cumulative effects on stability were observed upon binding of PI, proposed as the functional lipid in the *Mycobacterium* membrane involved in MscL mechanosensitivity<sup>9</sup>. Similarly, AqpZ was found to be non-selective for a number of lipids suggesting comparable binding modes, in line with the two-dimensional crystals of Aqp0<sup>13,27</sup>. The only exception to this non-selective lipid binding is CDL, which significantly stabilizes AqpZ and directly modulates function, as suggested in a previous *in vivo* study<sup>28</sup>. Interestingly, CDL levels in *E. coli* depend on growth conditions, a higher abundance of CDL being present in the stationary phase<sup>24</sup>, implying AqpZ function is fine-tuned through modulation of



**Figure 3 | Lipid binding to AmtB results in a range of stabilizing effects.** **a**, Ion mobility mass spectra of AmtB(PG)<sub>0–4</sub> molecules at 175 V with IM-MS unfolding data as described in Fig. 1b. Arrival times for the various charge states reveal non-native, trimeric structures. **b**, Stabilization of AmtB (+15) bound to different lipids and AmtB<sup>N72A/N79A</sup> (+15), a mutant form designed to disrupt the specific PG-lipid binding site. Gas-phase unfolding reveals that AmtB is weakly stabilized by most lipids but CDL and PG confer significant stabilization. AmtB<sup>N72A/N79A</sup> shows a reduction in stabilization by PG compared with the wild-type protein. Results are shown as in Fig. 1d with \* $P < 0.05$ .



**Figure 4 | Crystal structure of AmtB bound to PG.** **a**, AmtB, in cartoon representation, is viewed from the periplasm and perpendicular to the membrane plane. Resolved PG molecules (spheres) are oriented with headgroups on the periplasmic side, in the outer leaflet of the inner membrane. **b**, PG molecules at subunit interface are shown with residue labels and hydrophobic residues in contact with lipid tails (stick representation). Hydrogen bonds (dashed lines) are formed between the phospho headgroup and K84, and a water bridge (aqua sphere). **c**, Conformational change of (yellow arrows) residues 70–81, induced by binding PG. Superposition of AmtB structure without lipid (PDB accession code: 1U7G, purple) bound to PG (green).



CDL levels. In contrast to the other two systems, our finding that AmtB responds only weakly to the majority of lipids is consistent with its extensive transmembrane structure and inherent stability devoid of the lipid bilayer<sup>29</sup>. However, its unexpectedly high selectivity for PG led to the first crystal structure of AmtB in complex with lipid, allowing effects on stability to be rationalized and further validated by mutation of residues forming the specific lipid-binding site.

By investigating a diverse set of membrane proteins and their lipid-binding properties we have extended many seminal studies<sup>1–8</sup> by demonstrating different degrees of selectivity in lipid binding with AmtB > AqpZ > MscL. Despite these differences in selectivity, in all cases the most stabilizing lipids have a direct influence on the structure and function of the membrane proteins studied, as proposed for PI binding to MscL<sup>9</sup> and demonstrated for CDL and PG binding to AqpZ and AmtB, respectively. As membrane proteins are intimately embedded in the bilayer, the relative lipid abundances, as well as differences in selectivity, can enforce the recruitment of a local lipid environment<sup>30</sup>, thus providing an elegant means for fine-tuning membrane protein structure and function.

## METHODS SUMMARY

Detailed methods are described in Methods. In brief, membrane proteins were expressed as TEV protease cleavable fusion proteins containing a terminal His-tag in *Escherichia coli*. Membrane proteins were extracted from purified membranes, followed by affinity chromatography, TEV protease cleavage and gel filtration chromatography. Purified membrane complexes and lipids were prepared for IM-MS as previously described<sup>16</sup>. Lipids were added to solutions containing membrane protein complexes and gas-phase unfolding was induced at different collision voltages with a step size of 5 V on a modified Waters Synapt G1 HDMS instrument containing a linear drift cell. IM-MS data were analysed with software developed in-house. AqpZ functional assays were performed as previously described<sup>23</sup> using commercial or in-house *E. coli* total polar lipid extracts. Crystals of AmtB in the detergent C<sub>8</sub>E<sub>4</sub> were grown in the presence of PG in hanging-drop plates. Single crystals were mounted and flash frozen. Diffraction data were collected at the Diamond Light Source beam line I04. The structure was solved by molecular replacement, followed by automated and manual model building. MD was performed on membrane proteins in POPC bilayers, and candidate models were filtered using CCS measurements.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 November 2013; accepted 28 April 2014.

1. Singer, S. J. & Nicolson, G. L. The fluid mosaic model of the structure of cell membranes. *Science* **175**, 720–731 (1972).
2. Cantor, R. S. The influence of membrane lateral pressures on simple geometric models of protein conformational equilibria. *Chem. Phys. Lipids* **101**, 45–56 (1999).
3. Hunte, C. & Richers, S. Lipids and membrane protein structures. *Curr. Opin. Struct. Biol.* **18**, 406–411 (2008).
4. Lee, A. G. Biological membranes: the importance of molecular detail. *Trends Biochem. Sci.* **36**, 493–500 (2011).
5. Sanders, C. R. & Mittendorf, K. F. Tolerance to changes in membrane lipid composition as a selected trait of membrane proteins. *Biochemistry* **50**, 7858–7867 (2011).
6. Contreras, F. X., Ernst, A. M., Wieland, F. & Brugger, B. Specificity of intramembrane protein–lipid interactions. *Cold Spring Harb. Perspect. Biol.* **3**, <http://dx.doi.org/10.1101/cshperspect.a004705> (2011).
7. Whitelegge, J. P. Integral membrane proteins and bilayer proteomics. *Anal. Chem.* **85**, 2558–2568 (2013).
8. Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. Atomic structure of a voltage-dependent K<sup>+</sup> channel in a lipid membrane-like environment. *Nature* **450**, 376–382 (2007).
9. Zhong, D. & Blount, P. Phosphatidylinositol is crucial for the mechanosensitivity of *Mycobacterium tuberculosis* MscL. *Biochemistry* **52**, 5415–5420 (2013).

10. Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T. & Rees, D. C. Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel. *Science* **282**, 2220–2226 (1998).
11. Haswell, E. S., Phillips, R. & Rees, D. C. Mechanosensitive channels: what can they do and how do they do it? *Structure* **19**, 1356–1369 (2011).
12. Savage, D. F., Egea, P. F., Robles-Colmenares, Y., O’Connell, J. D. III & Stroud, R. M. Architecture and selectivity in aquaporins: 2.5 Å X-ray structure of aquaporin Z. *PLoS Biol.* **1**, e72 (2003).
13. Gonen, T. *et al.* Lipid–protein interactions in double-layered two-dimensional AQP0 crystals. *Nature* **438**, 633–638 (2005).
14. Khademi, S. *et al.* Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 Å. *Science* **305**, 1587–1594 (2004).
15. Barrera, N. P., Di Bartolo, N., Booth, P. J. & Robinson, C. V. Micelles protect membrane complexes from solution to vacuum. *Science* **321**, 243–246 (2008).
16. Laganowsky, A., Reading, E., Hopper, J. T. & Robinson, C. V. Mass spectrometry of intact membrane protein complexes. *Nature Protocols* **8**, 639–651 (2013).
17. Ruotolo, B. T. *et al.* Evidence for macromolecular protein rings in the absence of bulk water. *Science* **310**, 1658–1661 (2005).
18. Baldwin, A. J. *et al.* The polydispersity of  $\alpha$ B-crystallin is rationalized by an interconverting polyhedral architecture. *Structure* **19**, 1855–1863 (2011).
19. Hyung, S. J., Robinson, C. V. & Ruotolo, B. T. Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes. *Chem. Biol.* **16**, 382–390 (2009).
20. Pace, C. N. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280 (1986).
21. Hong, H., Joh, N. H., Bowie, J. U. & Tamm, L. K. in *Methods Enzymol.* Vol. 455 (eds Michael L. Johnson, Jo M. Holt, & Gary K. Ackers) 213–236 (Academic Press, 2009).
22. Powl, A. M., East, J. M. & Lee, A. G. Lipid–protein interactions studied by introduction of a tryptophan residue: the mechanosensitive channel MscL. *Biochemistry* **42**, 14306–14317 (2003).
23. Borgnia, M. J., Kozono, D., Calamita, G., Maloney, P. C. & Agre, P. Functional reconstitution and characterization of AqpZ, the *E. coli* water channel protein. *J. Mol. Biol.* **291**, 1169–1179 (1999).
24. Tan, B. K. *et al.* Discovery of a cardiolipin synthase utilizing phosphatidylethanolamine and phosphatidylglycerol as substrates. *Proc. Natl Acad. Sci. USA* **109**, 16504–16509 (2012).
25. Yau, W. M., Wimley, W. C., Gawrisch, K. & White, S. H. The preference of tryptophan for membrane interfaces. *Biochemistry* **37**, 14713–14718 (1998).
26. Iscla, I. & Blount, P. Sensing and responding to membrane tension: the bacterial MscL channel as a model system. *Biophys. J.* **103**, 169–174 (2012).
27. Hite, R. K., Li, Z. & Walz, T. Principles of membrane protein interactions with annular lipids deduced from aquaporin-0 2D crystals. *EMBO J.* **29**, 1652–1658 (2010).
28. Romantsov, T., Battle, A. R., Hendel, J. L., Martinac, B. & Wood, J. M. Protein localization in *Escherichia coli* cells: comparison of the cytoplasmic membrane proteins ProP, LacY, ProW, AqpZ, MscS, and MscL. *J. Bacteriol.* **192**, 912–924 (2010).
29. Blakey, D. *et al.* Purification of the *Escherichia coli* ammonium transporter AmtB reveals a trimeric stoichiometry. *Biochem. J.* **364**, 527–535 (2002).
30. Engelman, D. M. Membranes are more mosaic than fluid. *Nature* **438**, 578–580 (2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. Hobman and D. Lee for providing gene-doctoring plasmids, Z. Guan and C. Li for providing *E. coli* strains and plasmids containing cardiolipin genes. We also thank D. Staunton and N. Housden for training on the stopped-flow apparatus; and T. Mize, J. Benesch, M. McDonough, T. Walton and D. Rees for discussions. We also gratefully acknowledge E. Lowe and S. Lea for organizing synchrotron proposals and Diamond Light Source beam line I04 and staff, the Medical Research Council (MRC), BBSRC and ERC advanced grant (IMPRESS) for funding. A.L. is a Nicholas Kurti Junior Research Fellow of Brasenose College, A.J.B. is a BBSRC David Phillips’ Fellow and C.V.R. is a Royal Society Professor.

**Author Contributions** A.L., E.R., and C.V.R. designed the research. A.L. and E.R. performed the experiments. T.M.A. assisted A.L. and E.R. in protein expression and purification. M.B.U. carried out molecular dynamics. M.T.D., A.J.B. and A.L. designed and performed post-molecular dynamics analyses. A.L., E.R., T.M.A. and A.J.B. developed IM-MS analysis software. A.L. and E.R. analysed the data. A.L., E.R. and C.V.R. wrote the paper with input from the other authors.

**Author Information** Atomic coordinates and structure factors for the crystal structure have been deposited with the Protein Data Bank (PDB) under accession code 4NH2. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.L. ([art.laganowsky@chem.ox.ac.uk](mailto:art.laganowsky@chem.ox.ac.uk)) or C.V.R. ([carol.robinson@chem.ox.ac.uk](mailto:carol.robinson@chem.ox.ac.uk)).

## METHODS

**Plasmid construction.** Two expression plasmids were constructed by subcloning the multiple cloning site region from XbaI (New England Biolabs) and BspI (New England Biolabs) of pET23b (Novagen) into the backbone pET15b (Novagen). The resulting engineered vector was linearized by NdeI (New England Biolabs) and XhoI (New England Biolabs), gel purified (QIAquick Gel Extraction Kit, Qiagen), and used in subsequent Infusion cloning reactions (Clontech) to generate a TEV protease cleavable C-terminal fusion to superfolder GFP (subcloned from Gandhi *et al.*<sup>31</sup>) followed by a 6× His-tag or a TEV protease cleavable N-terminal fusion to maltose binding protein (MBP) preceded by a secretion signal peptide (pelB) and 10× His-tag subcloned from Hilf *et al.*<sup>32</sup>. The resulting N- and C-terminal fusion vectors were linearized with NheI and XhoI or NdeI and NheI, gel purified, and used in subsequent Infusion cloning reactions. AmtB (residues 26–428), AqpZ and MscL genes were amplified by polymerase chain reaction (PCR) with Phusion high-fidelity DNA polymerase (New England Biolabs) from prepared *E. coli* BL21 (DE3) genomic DNA (Qiagen) or template plasmid DNA with primers designed for an Infusion cloning reaction using the manufacturer's online tool. The PCR products were purified by agarose gel electrophoresis and extracted using the QIAquick gel extraction kit (Qiagen). The PCR products and linearized vectors (described above) were used in Infusion cloning reactions (Clontech) to generate MBP–AmtB, AqpZ–GFP and MscL–GFP. MBP–AmtB<sup>N72A/N79A</sup> was achieved by two-rounds of site-directed mutagenesis using a QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) according to the manufacturer's protocol. AqpZ harbouring a TEV protease cleavable N-terminal 6× His-tag (NHIS–AqpZ) was constructed through an infusion cloning reaction of AqpZ and a custom pRSDuet-1 vector (Novagen), the N-terminal coding sequence between NcoI and BamHI was replaced with the following nucleotide sequence: CCATGGGCAGCAGCCATCA CCATCATCACCACGAGAACCTGTACTTCCAGGGTGGATCC, linearized with BamHI and XhoI. All constructs were verified by DNA sequencing.

**Generation of *E. coli* BL21(DE3) Δ*mscL*::*Kan*<sup>R</sup> strain.** Homologous recombination mediated chromosomal disruption of *mscL* was performed using the Gene Doctoring method<sup>33</sup>. Briefly, 900 base pairs flanking the 5' and 3' ends of the *mscL* gene were subcloned into the pDOC-K plasmid resulting in these regions flanking a kanamycin resistance (*Kan*<sup>R</sup>) cassette. This plasmid and the recombinering plasmid, pACBSCE, were transformed into *E. coli* BL21 (DE3) Gold (Agilent). Homologous recombination was carried out by induction of λ-Red recombinase system and ScaI endonuclease with arabinose. Recombinants harbouring gene disruption with *Kan*<sup>R</sup> were selected on kanamycin and sucrose, and verified by colony PCR. Calcium chloride competent cells were made as described by Drew *et al.*<sup>34</sup>.

**Membrane protein expression.** MBP–AmtB and AqpZ–GFP plasmids were transformed into *E. coli* BL21 (DE3) Gold (Agilent). MBP–AmtB<sup>N72A/N79A</sup> and NHIS–AqpZ were transformed into *E. coli* OverExpress C43 (DE3) (Lucigen). MscL–GFP was transformed into *E. coli* BL21(DE3) Δ*mscL*::*Kan*<sup>R</sup> (described above). Several colonies were inoculated into 50 ml LB Miller (5 g yeast extract, 10 g peptone from casein, and 10 g sodium chloride per litre) and grown overnight at 37 °C. One litre of LB in 2 litre shaker flasks was inoculated with 7 ml of overnight culture and grown at 37 °C until the culture reached OD<sub>600nm</sub> (OD<sub>600</sub>) between 0.6 to 0.8. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the culture at a final concentration of 0.5 mM and grown for 3 h at 37 °C. Cells were collected by centrifugation at 5,000g for 10 min at 4 °C. Cell pellets were resuspended in phosphate-buffered saline, pelleted by centrifugation at 5,000g for 10 min at 4 °C and stored at –80 °C.

**Purification of membranes.** Membranes were purified as described by Newby *et al.*<sup>35</sup> with minor modifications. Cell pellets were thawed and resuspended at 20 ml per litre of culture in Buffer A (300 mM sodium chloride, and 20 mM 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris), pH 7.4 at room temperature) supplemented with a complete protease inhibitor tablet (Roche) and 5 mM beta-mercaptoethanol (BME). The cell suspension was passed several times through an M-110 PS microfluidizer (Microfluidics) at 19,000 psi. Insoluble material was pelleted by centrifugation at 20,000g for 25 min at 4 °C. Membranes were pelleted by centrifugation at 100,000g for 2 h at 4 °C. Membranes were resuspended in ice-cold buffer B (100 mM sodium chloride, 20% glycerol, 5 mM BME and 20 mM Tris, pH 7.4 at room temperature), homogenized using a Potter–Elvehjem Teflon pestle and glass tube. For MscL–GFP, glycerol was omitted from Buffer B. Resuspended membranes were used either directly or flash frozen in liquid nitrogen and stored at –80 °C.

**AmtB–GFP detergent screen.** A membrane protein detergent screen on AmtB–GFP was performed as previously described<sup>16</sup> with minor modifications. Briefly, AmtB–GFP was extracted from purified membranes (described above) with 1–2% (w/v) of the detergent of interest in buffer B and incubated for one to three hours at room temperature or overnight at 4 °C with gentle agitation. Insoluble material was pelleted by centrifugation at 20,000g for 25 min at 4 °C. The clarified supernatant was loaded onto small Ni–NTA agarose (Qiagen) drip columns (Bio-spin Chromatography columns, Bio-Rad) and washed with several column volumes of buffer C (200 mM sodium chloride, 20 mM imidazole, 5 mM BME, 50 mM Tris, pH 8.0 at

room temperature) supplemented with two times the critical micelle concentration (CMC) of the detergent of interest. AmtB–GFP was eluted with two column volumes of buffer D (10 mM sodium chloride, 250 mM imidazole, 5 mM BME, 2× CMC detergent of interest, and 50 mM Tris, pH 8.0 at room temperature). Eluted protein was concentrated using a 100 kDa Molecular Weight Cutoff (MWCO) concentrator and buffer exchanged into MS Buffer (2× CMC detergent of interest and 200 mM ammonium acetate, pH 8.0 with ammonium hydroxide) using a centrifugal buffer exchange device (Micro Bio-Spin 6, Bio-Rad).

**Purification of membrane proteins.** MBP–AmtB, MBP–AmtB<sup>N72A/N79A</sup> and AqpZ–GFP were extracted from purified membranes in buffer B supplemented with 200 mM OG and incubated with gentle agitation overnight at 4 °C. MscL was extracted from purified membranes with glycerol-free buffer B supplemented with 1% OGNG. Extracted membrane proteins were clarified by centrifugation at 20,000g for 25 min at 4 °C. Supernatant was filtered before loading onto a 5 ml HisTrap–HP column (GE Healthcare, Piscataway, NJ) equilibrated in buffer E (200 mM sodium chloride, 10% glycerol, 20 mM imidazole, 0.025% DDM and 50 mM Tris, pH 7.4 at room temperature). After the clarified supernatant was loaded, the column was initially washed with 40–50 ml of DDM-free buffer E supplemented with 1% OG for MBP–AmtB and AqpZ–GFP and 0.5% OGNG for MscL–GFP. Membrane proteins were then exchanged into several column volumes of buffer E until a steady baseline was reached. Membrane protein fusions were eluted with a linear gradient to 100% in two column volumes of buffer F (100 mM sodium chloride, 10% glycerol, 500 mM imidazole, 0.025% DDM and 50 mM Tris, pH 7.4 at room temperature). Peak fractions were pooled, supplemented with 5 mM BME and His-tagged TEV protease<sup>36,37</sup>, and dialysed against buffer G (150 mM sodium chloride, 10% glycerol, 20 mM imidazole, 0.025% DDM and 50 mM Tris, pH 7.4 at room temperature) overnight at 4 °C. After overnight incubation, samples were filtered and passed back over a 5 ml HisTrap–HP column equilibrated in buffer E. Flow-through containing the untagged membrane protein was collected and concentrated using a 100 kDa MWCO concentrator. Notably, tag-removed membrane proteins with fusions to the C terminus contained the additional protein sequence ASGENLYFQ, or GAS for N-terminally tagged proteins, resulting from the TEV protease recognition sequence and cloning restriction site. Concentrated protein was either used immediately or flash-frozen in liquid nitrogen and stored at –80 °C. Protein concentration was measured using Biomat UV detector with the following calculated extinction coefficients<sup>38</sup> of 1.57, 1.484, and 0.265 M<sup>–1</sup> cm<sup>–1</sup> for AmtB, AqpZ and MscL, respectively.

NHIS–AqpZ was purified as previously described<sup>39</sup>. Briefly, extracted and clarified membrane proteins were loaded onto a 5 ml HisTrap–HP column (GE Healthcare, Piscataway, NJ) equilibrated in buffer (200 mM sodium chloride, 10% glycerol, 20 mM imidazole, 1% OG and 50 mM Tris, pH 7.4 at room temperature). NHIS–AqpZ was eluted with a linear gradient to 100% elution buffer (100 mM sodium chloride, 10% glycerol, 500 mM imidazole, 1% OG and 50 mM Tris, pH 7.4 at room temperature) over two column volumes. Peak fractions containing NHIS–AqpZ were concentrated before imidazole removal using two 5 ml HiTrap desalting columns (GE Healthcare, Piscataway, NJ) in tandem. Peak fractions were pooled and NHIS–AqpZ was concentrated using a 50 kDa MWCO concentrator.

**Preparation of membrane proteins for native mass spectrometry.** Either flash-frozen samples thawed on ice or fresh samples were detergent exchanged by gel filtration chromatography. Membrane protein samples were injected onto a Superdex 200 GL 10/300 (GE Healthcare) column equilibrated in buffer H (130 mM sodium chloride, 10% glycerol and 50 mM Tris, pH 7.4 at room temperature) supplemented with either 0.5% of C<sub>8</sub>E<sub>4</sub>, 0.4% of NG, 0.116% OGNG or 0.025% of DDM. Peak fractions containing the detergent-exchanged membrane protein complex were concentrated. A 50 kDa MWCO concentrator was used to concentrate samples in the detergent C<sub>8</sub>E<sub>4</sub> and 100 kDa MWCO concentrator used on membrane protein complexes in all other detergents. Concentrated proteins were either used directly or flash-frozen in liquid nitrogen and stored at –80 °C. Notably, we found no observable difference in mass spectra quality after a single freeze-thaw of membrane proteins throughout our purification regime. Purified membrane proteins were buffer exchanged into MS Buffer (two times the CMC of detergent of interest and 200 mM ammonium acetate, pH 7.2–8.0 with ammonium hydroxide) using a centrifugal buffer exchange device (Micro Bio-Spin 6, Bio-Rad) as previously described<sup>16</sup>. Membrane proteins detergent exchanged into MS buffer supplemented with C<sub>8</sub>E<sub>4</sub> or OGNG could be flash frozen without compromising mass spectra quality unlike MS Buffer supplemented with NG or DDM.

**Preparation and titration of phospholipids.** Phospholipids were purchased from Avanti (Avanti Polar Lipids Inc., Alabama, USA) and prepared at stock concentrations of 10 mg ml<sup>–1</sup> in 200 mM ammonium acetate pH 8.0 as previously described<sup>16</sup>. The membrane protein complex to phospholipid to detergent ratio (P:L:D) was optimized for each membrane protein complex to achieve consistency of nanoelectrospray and resolved mass spectral peaks of bound phospholipid throughout the gas-phase unfolding series (Extended Data Fig. 2b). Phospholipid preparations were added to buffer exchanged membrane protein complexes followed by

equilibration at room temperature for 10 to 30 min before mass spectrometry analysis. Notably, by mass spectrometry we did not observe a significant increase in bound phospholipid with longer incubation times.

**Mass spectrometry.** Mass spectrometer settings were initially set to values for membrane proteins as previously described<sup>16</sup> for a modified Q-TOF 2 mass spectrometer (Micromass, Manchester, UK) with a Z-spray source<sup>40</sup>. Typical instrument values were 5–7  $\mu$ bar for source pressure, 1.5–1.8 kV for capillary voltage, 150–190 V for cone voltage, 1–10 V for extraction voltage, 180–200 V for collision voltage, argon for collisional gas and 0.2–0.3 MPa argon gas pressure.

**Ion mobility mass spectrometry.** Ion mobility measurements were performed on a Synapt G2 (where noted) or on a modified Synapt G1 HDMS instrument with the travelling-wave ion mobility cell replaced by an 18-cm drift cell with radial RF ion confinement and a linear voltage gradient to direct ions along the axis of transmission to the time-of-flight (TOF) mass analyser<sup>41</sup>. The ion mobility mass spectrometer was typically set to a source pressure of 5–7 mBar, capillary voltage of 1.4–1.7 kV, capillary nanoflow of 0.03–0.2 mBar and argon as collision gas with flow rate set to 5–8 ml min<sup>-1</sup> ( $\sim 6.6$ – $6.7$  e<sup>-2</sup> mBar). Collision voltage ranged from 50 to 240 V with measurements taken at 5 V steps for monitoring gas-phase unfolding of membrane protein complexes. The sample and extraction cone, and trap bias voltages, quadrupole profile, and collision (trap) gas pressure were optimized for maximal ion intensity of the target membrane protein complex. Helium was the drift cell gas and was set to a pressure of  $\sim 1.6$ – $1.8$  Torr (40 ml min<sup>-1</sup> flow rate). Pusher time, pulse width and TDC inhibit were set to 180, 7 and 7  $\mu$ s, respectively. The ion guide/source (300 ms<sup>-1</sup>, 10 V), trap (300 ms<sup>-1</sup>, 0.2 V) and transfer (100 ms<sup>-1</sup>, 10 V) travelling wave velocities and wave heights were kept constant. The mobility release time was set at 200  $\mu$ s with a trap voltage of 30 V and extract voltage of 10 V. Pressure and temperature of the ion mobility cell was measured directly.

Ion drift time ( $t_d$ ) was determined as previously described<sup>41</sup>. Briefly,  $t_d$  was determined by subtracting the non-drift time component ( $t_o$ ) from the ion arrival time ( $t_a$ )

$$t_d = t_a - t_o$$

$t_o$  was determined by extrapolation from linear plots of  $t_a$  versus  $V^{-1}$  over ten different voltage potentials (Extended Data Fig. 1d). Ion drift time was converted to a rotationally averaged ion-neutral collision cross section (CCS) ( $\Omega$ ) using the Mason-Schamp equation<sup>42</sup>

$$\Omega = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{ze}{(k_b T)^{\frac{1}{2}} N} \frac{1}{L^2} \frac{t_d V p_o}{p} \frac{T}{T_o} \sqrt{\frac{1}{\mu}}$$

here  $z$  is the charge on the ion,  $\mu$  the reduced mass of the ion-neutral in the drift chamber,  $N$  is the drift gas number density,  $k_b$  is the Boltzmann constant,  $e$  is the charge on the electron,  $V$  is the potential voltage across the drift cell,  $p$  is the pressure of the drift cell,  $T$  is the thermodynamic temperature of the drift cell,  $L$  is the length of the drift tube,  $T_o$  is standard temperature and  $p_o$  is standard pressure.

Travelling wave ion mobility measurements (TW-IMS) were performed on a commercially available Synapt G2 (Waters). Source pressure set to 5–7 mBar, capillary voltage of 1.4–1.7 kV, capillary nanoflow of 0.03–0.2 mBar and argon as collision (trap) gas at flow rate of 8.0 ml min<sup>-1</sup>. Collision voltage ranged from 50 to 200 V with measurements taken at 5 V step intervals for monitoring gas-phase unfolding of membrane complexes. The sample and extraction cone and trap bias voltages, quadrupole profile, and collision gas pressure were optimized for maximal ion intensity of the target membrane protein complex. Helium was the IMS entrance cell gas set to a flow rate of 180 ml min<sup>-1</sup> and the drift cell gas was nitrogen set to flow rate of 90 ml min<sup>-1</sup>. The EDC delay coefficient was 1.57  $\mu$ s. The trap wave velocity and wave height were 300 ms<sup>-1</sup> and 8.0 V, respectively. The IMS wave velocity and wave height were 300 ms<sup>-1</sup> and 30.0 V, respectively. The transfer wave velocity and wave height were 100 ms<sup>-1</sup> and 2.0 V, respectively. The mobility release time was set at 200  $\mu$ s with a trap voltage of 30 V and extract voltage of 10 V. Gas-phase unfolding profiles are largely similar to those collected on a drift cell instrument (Extended Data Fig. 6d).

**Collision cross-section calculations.** The projection approximation (PA) calculation within a modified version of MOBCAL<sup>18,43,44</sup> used to calculate the  $\Omega_{PA}$  from respective crystal structure coordinates. CCS values were corrected  $\Omega_{Corr}$  by the correction factor of 1.14 and missing residues between the protein construct ( $M_{Exp}$ ) and crystal structure ( $M_{PDB}$ )<sup>45,46</sup>

$$\Omega_{Corr} = 1.14 \Omega_{PA} \left( \frac{M_{Exp}}{M_{PDB}} \right)^{\frac{2}{3}}$$

This correction has been shown to give values that correlate with known values to  $\pm 3\%$ <sup>41,47</sup>.

**Gas-phase unfolding data analysis and modelling.** Mass spectra and ion mobility data were analysed using in-house software written in the Python programming

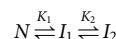
language with a graphical user interface constructed using wxPython (manuscript in preparation). Briefly, MS-only data collected on a Q-TOF 2 instrument was smoothed with MassLynx software before being imported and linearized at a step size of one in  $m/z$ . Theoretical mass spectra were predicted following equations described in Stengel *et al.*<sup>48</sup>. Mass, average charge state, width of charge state distribution, resolution, and concentration were fitted to experimental data with least squares regression of the pseudo- $\chi^2$  function<sup>48</sup>.

Ion mobility data were imported using a custom script provided by Waters at an  $m/z$  resolution of four in drift data and one in mass spectra for fitting. Mass spectra were smoothed and theoretical mass spectra were fitted as described above. A standardized integration window was employed for extracting arrival time distributions (ATD) using a resolution ( $R = m/z \div \Delta m/z$ ) of 1,000 and a height cut-off of 25% for the modelled Gaussian to the respective  $m/z$  peak. To generate gas-phase unfolding plots, ATDs were extracted for the given  $m/z$  peak using the standardized integration window followed by normalization. To convert to a CCS axis, arrival time was directly converted to CCS using equations above. This procedure generates gas-phase unfolding data and are plotted using Matplotlib<sup>49</sup>. All ion mobility mass spectra are shown in linear scale.

To a first order approximation, gas-phase unfolding of protein complexes, like unfolding in solution, will be linearly dependent ( $m$ ) on collision voltage (CV)

$$\Delta G = \Delta G^0 - m[CV]$$

where  $\Delta G$  is free energy of unfolding and  $\Delta G^0$  is the free energy of the system at zero collision voltage in the gas-phase. Using this assumption, two, three, four and five state linear equilibrium unfolding models were used to model the abundance of native and non-native species. Notably, two new free parameters are introduced for each new species. Two, three, four and five state unfolding models were fitted to the data, and an  $F$ -test was used to decide if the fitting quality is sufficiently improved as to justify the inclusion of additional fitting parameters. The most abundant charge state of the three membrane protein complexes agreed with a four state unfolding model. For example, in the case of three unfolding species we used the following thermodynamic denaturing unfolding model<sup>20</sup>.



Expansion of  $\Delta G_{N \rightleftharpoons I_1}$  gives the following,

$$\Delta G_{N \rightleftharpoons I_1} = \Delta G_{N \rightleftharpoons I_1}^0 - m_{N \rightleftharpoons I_1} [CV]$$

$$\Delta G_{N \rightleftharpoons I_1} = -RT \ln K_1$$

$$K_1 = \frac{F_{I_1}}{F_N}$$

where  $F$  is a mole fraction. The sum of mole fractions of native ( $F_N$ ) and non-native species ( $F_{I_1}$  and  $F_{I_2}$ ) is equal to 1

$$F_N + F_{I_1} + F_{I_2} = 1.$$

We can generalize this taking into account the entire relationships one can obtain the fraction of native and non-native given  $K_1$  and  $K_2$

$$F_N = \frac{1}{1 + K_1 + (K_1 K_2)}$$

$$F_{I_1} = \frac{K_1}{1 + K_1 + (K_1 K_2)}$$

$$F_{I_2} = \frac{(K_1 K_2)}{1 + K_1 + (K_1 K_2)}$$

The respective  $\Delta G^0$  and  $m$  parameters were solved numerically. In the two species case one can obtain a closed form solution, however, it is both efficient and general to solve them numerically.

An irreversible kinetic unfolding model as an alternative for modelling gas-phase unfolding data was also explored. The irreversible unfolding model defines the degree of unfolding as being related to time. By changing the potential across the drift cell, the time ions spend in the ion mobility drift cell post collision-induced unfolding can be altered over several orders of magnitude. The relative abundances of unfolded intermediates species as a function of drift time was monitored for a membrane protein complex. There was no considerable change in relative abundances over a fourfold increase in time (Extended Data Fig. 6e, f) demonstrating that the unfolding mechanism does not agree with an irreversible kinetic model for the unfolding model and no further unfolding occurs post collision cell. Data are in excellent agreement however with a reversible model as described above.



Ion mobility intensities were extracted for native and unfolded CCS species at various collision voltages. With the assumption that signal intensity is approximately proportional to species concentration, the data can be converted to a molar fraction to generate two dimensional (2D) data, which can be fitted to an equilibrium unfolding model. Minimized 2D parameters were used as a seed for three dimensional (3D) fitting of gas-phase unfolding data.

From the unfolding curve, and parameters  $m$ ,  $\Delta G^\circ$ ,  $CCS_\mu$  (mean of CCS species),  $CCS_\sigma$  (standard deviation of CCS species), the 3D data can be fit. First, a CCS range was selected for each native and unfolded species in the unfolding plot. These CCS ranges provided the initial parameters used to model various CCS species identified by inspection. Next, to make the calculations more efficient, the gas-phase unfolding data were trimmed to a minimum and maximum value derived by either subtraction or addition of 10% times the native and last unfolding CCS species, respectively. Trimming resulted in gas-phase unfolding containing on average around 900 data points. The equilibrium unfolding model described above was seeded with the parameters from the minimized 2D data and the molar fraction of species was normalized to match the normalization of experimental data. The fitting surface was rugged and required a sophisticated minimization procedure to maximize the chance of finding global minimum. First the 3D model was minimized using quasi-Newton method of Broyden, Fletcher, Goldfarb, and Shanno (BFGS) available in Scipy<sup>50</sup>. This provided the boundaries for bio-inspired algorithms; all parameters were set to boundaries of  $\pm 25\%$  with the exception of the CCS species peak centres which were set to  $\pm 5\%$ . The 3D model was minimized using the modified differential algorithm (de\_1220) available in PyGMO<sup>51,52</sup> using the following parameters: population of 20; 12 evolutions; 250 generations; and 8 islands. Other algorithms, such as particle swarm and bee colony were screened however the modified differential algorithm produced similar or higher  $R^2$  values in considerably less computation time.

The resulting minimized 3D model enabled quantification of unfolding transitions and stabilization. An unfolding transition occurs when a species with a determined CCS transitions to another CCS species, for example see Extended Data Fig. 6c. The transition midpoint,  $CV_{50}$  is determined by:

$$CV_{50,i} = \frac{\Delta G^\circ}{m}$$

where  $CV_{50,i}$  is the value at which 50% of a specific transition state  $i$  is depleted. This metric was found to be particularly useful for this application as it removes correlations between  $\Delta G^\circ$  and  $m$  inherent in the fitting procedure. The stabilization of a protein due to the presence of additional phospholipids was quantified according to:

$$\Delta CV_{50,i} = CV_{50,i}^{\text{lipid}} - CV_{50,i}^{\text{apo}}$$

This comparative approach takes into account to first order any variability in transition collision voltages between repeated measurements for a given protein.

To compare between various proteins and different unfolding states, as well as to average out any systematic and/or inherent variability in droplet formation, the average stabilization was calculated for the bound ligand by

$$\text{Stabilization} = \frac{z \sum_i^T \Delta CV_{50,i}}{\zeta}$$

where  $\zeta$  is the number of transitions,  $i$  is a specific transition, and  $z$  is the charge state. Averages and standard error of the means was calculated for three repeated measurements. Statistical significance between means were determined by a one-way ANOVA test ( $n = 3$ ) followed by post-hoc Tukey analysis. A Shapiro–Wilk test (at the 0.05 level) was used to infer the data was drawn from a normally distributed population. A Brown–Forsythe test (at the 0.05 level) was performed to infer homogeneity of variance before means comparison. All statistical tests were performed using OriginPro 8.5.1.

**Molecular dynamics preparation and simulation.** Crystal structures for the Ammonia Channel AmtB (1U7G)<sup>14</sup>, Aquaporin Z (1RC2)<sup>12</sup> and MscL (2OAR)<sup>10</sup> were obtained from the Protein Data Bank and prepared. Molecular simulations were performed using GROMACS 4.5 (<http://www.gromacs.org>)<sup>53</sup>, with the OPLS all-atom protein force field<sup>54</sup>, OPLS united-atom lipid parameters for palmitoylcholine (POPC)<sup>55</sup> and the TIP3P water model<sup>56</sup>. Initial configurations were created using hippo beta (<http://www.biowerkzeug.com>).

Electrostatic interactions were computed using the particle mesh Ewald (PME) method<sup>57</sup>, and a cut-off of 10 Å was used for van der Waals interactions. Bonds involving hydrogen atoms were restrained using LINCS<sup>58</sup>. Simulations were run with a 2 fs integration time step and neighbour lists were updated every five steps. All simulations are performed in an NPT ensemble, without additional applied surface tension. Solvent (water, ions), lipids, and the protein were each coupled separately to a heat bath at a temperature of 30 °C with time constant  $\tau_T = 0.1$  ps using weak temperature coupling<sup>59</sup>. Atmospheric pressure of 1 bar was maintained

using weak semi-isotropic pressure coupling with compressibility  $\kappa_z = \kappa_{xy} = 4.6 \times 10^{-5} \text{ bar}^{-1}$  and time constant  $\tau_P = 1 \text{ ps}^{60}$ . A complete list of simulations is given in Extended Data Fig. 4b.

Each protein phospholipid pair generated by molecular dynamics (MD) was treated as a candidate model. For example,  $\sim 206,000$  such models were generated for AqpZ and filtered by comparison of calculated and experimental CCS values. More specifically, within each MD snapshot there are  $\sim 80$  lipids that could be in accord with the CCS measurements for one lipid bound to AqpZ. After filtering by agreement with CCS, an ensemble of protein phospholipid complexes consistent with our data was identified. As the MD is used to provide a source of potential lipid associated structures, the results are largely invariant to the choice of force field and lipid molecule.

#### Identifying lipid binding sites from MD simulations and CCS measurements.

The object is to determine which MD frames are in best agreement with the experimental CCS data. For every simulated system, one frame was extracted every 0.2 ns generating 1,000 PDB files for a 500 ns simulation time. For each extracted frame, all lipids within 6 Å of the protein were identified. All possible combinations of these two lipids and protein were selected. For example for AqpZ and one lipid typically yielded 80 extracted combinations per frame (Extended Data Fig. 4c). The CCS value of each combination per frame ( $f$ ) was calculated using a modified version of MOBCAL<sup>18,43,44</sup>. A ratio ( $R_{f,\text{calc}}$ ) was determined using the calculated CCS values for each of the +1 lipid and the +2 lipid-bound states

$$R_{f,l,\text{calc}} = \frac{CCS_{\text{protein+lipid},f,l}}{CCS_{\text{protein},f}}$$

where  $CCS_{\text{protein+lipid}}$  is CCS of protein plus number of lipid(s) ( $l$ ) and  $CCS_{\text{protein}}$  is CCS of protein. Using a ratio avoided any potential issues associated with taking the difference of uncorrected CCS values determined experimentally. These measurements enable the calculation of three chi-squared ( $\chi^2$ ) values for a given combination of two lipids and protein

$$\chi_{f,l}^2 = \frac{(R_{f,l,\text{calc}} - R_{f,l,\text{exp}})^2}{\sigma_l^2}$$

where  $R_{f,\text{exp}}$  and  $\sigma$  are the average and standard deviation of the experimental CCS ratio for data with one or two lipid molecules bound from three repeated measurements, and  $R_{f,\text{exp}}$  is the CCS ratio derived from theoretical calculations. The overall chi-squared values were calculated from:

$$\chi_f^2 = \chi_{f,2}^2 + \chi_{f,1a}^2 + \chi_{f,1b}^2$$

where  $\chi_{f,2}^2$  is the chi-squared calculated for two lipids bound, and  $\chi_{f,1a}^2$  and  $\chi_{f,1b}^2$  are the chi-squared values for each individual lipid molecule with the protein. In the case where only the +1 lipid CCS data are analysed, neglecting the +2 lipid data, only one  $\chi^2$  value contributed towards the sum. The probability ( $P_f$ ) that a given protein phospholipid pair with protein matches the data are given by the following Bayesian probability:

$$P_f \propto e^{-\frac{(\chi^2 - \chi_{\min}^2)}{2}}$$

where  $\chi_{\min}^2$  is the minimum chi-squared value obtained across the entire simulation, whose corresponding  $P_f$  is equal to 1, representing the most probable structure. It follows that all candidate lipid pair and protein combinations with a  $P_f$  value greater than 0.785 will be one half standard deviation from the most probable structure. The ensemble of structures identified with  $P_f > 0.785$  can be considered the most probable arrangement of lipids and protein found to be consistent with the experimental data. This ensemble of structures was projected onto the surface of the protein to identify the most probable locations of contact between the protein and the lipid.

**Total polar lipid extracts from cardiolipin-deficient *E. coli* strain, BKT22.** The cardiolipin-deficient *E. coli* strain<sup>24</sup>, BKT22 ( $\Delta clcA$ ,  $\Delta clcB$ ,  $\Delta clcC$ ,  $\Delta ymdB::Kan^R$ ) and calcium chloride competent BKT22 cells were transformed with pBAD-YC (plasmid containing *clsC* and *ymdB* genes<sup>24</sup>), referred herein as BKT22-YC, and grown overnight in LB media. Overnight cultures were diluted to an OD<sub>600</sub> of 0.03 in LB media with the exception that arabinose was added to a final concentration of 0.2% for BKT22-YC to induce expression of cardiolipin genes, *clsC* and *ymdB*. The strains were grown for six hours to stationary phase (OD<sub>600</sub>  $\sim 2.0$ ). Cells were collected by centrifugation at 3,500g at 4 °C for 10 min. Cell pellets were resuspended in phosphate-buffered saline followed by centrifugation.

Lipids were extracted using the Folch method<sup>61</sup>. Briefly, cells were resuspended in 20 ml of 2:1 (v/v) chloroform:methanol per gram of cells and agitated on a stir plate for 20 min. The mixture was then clarified using filter paper and washed with 0.2 volume of water. The extract was briefly vortexed before centrifugation at 800g

to separate phases. The upper aqueous phase was removed and the interface was washed with 1:1 (v/v) methanol:water without disturbing the lower organic phase. The lower organic phase was recovered and evaporated under a stream of nitrogen. The dried total lipid extract was then resuspended in diethyl ether to extract polar lipids and evaporated under nitrogen then under vacuum overnight.

**Thin-layer chromatography.** HPTLC Silica gel 60 aluminium plates (Merck Millipore) were pre-developed with 1:1 (v/v) chloroform:methanol and left to air-dry overnight. The plates were then impregnated with an adsorbent modifier (1:1 (v/v) 1.2% boric acid in absolute ethanol-water) followed by activation at 100 °C for one hour<sup>62</sup>. Samples were spotted and plates developed in 65:25:5 (v/v) chloroform:methanol:acetic acid. After air-drying for 30 min, plates were sprayed with 10% sulphuric acid in ethanol and charred at 200 °C (ref. 63). Molar per cent (mol %) of lipid species was determined from the intensity of lipid species using ImageJ 1.47v software<sup>64</sup>.

**Proteoliposome/liposome preparation.** *E. coli* total polar lipid extract (acetone/ether preparation; Avanti Polar Lipids) was dissolved at 20 mg ml<sup>-1</sup> in 2:1 (v/v) chloroform:methanol to obtain a clear solution. Solvent was evaporated under a stream of nitrogen gas then under vacuum overnight. *E. coli* total polar lipid (EPL), BKT22 and BKT22-YC extracts were hydrated in 2 mM BME to a final concentration of 50 mg ml<sup>-1</sup> and left to incubate for 1 h at room temperature with constant agitation. These were then split into aliquots and frozen. Lipids were always handled under a nitrogen atmosphere.

Liposomes and proteoliposomes were prepared following previously established protocols<sup>23,39</sup>. Briefly, lipid stocks were diluted into borosilicate tubes to a final concentration of 45 mg ml<sup>-1</sup> in 100 mM MOPS-Na pH 7.5 and pulsed in a bath sonicator until a clear suspension was obtained. A reconstitution mixture (400 µl) was prepared in a borosilicate tube at room temperature by sequentially adding 100 mM Mops-Na (pH 7.5), 1.25% (wt/vol) OG, purified NHis-AqpZ for proteoliposomes (final concentration 100 µg ml<sup>-1</sup>), and 10 mg ml<sup>-1</sup> sonicated lipids. After incubation for 1 h at room temperature, the mixture was diluted 25-fold into 20 mM HEPES pH 7.5. The liposomes were collected by centrifugation 140,000g for one hour and resuspended to a final volume of 1.5 ml in 20 mM HEPES pH 7.5. Liposomes and proteoliposomes were stored on ice before measurement. EPL, BKT22 and BKT22-YC liposomes/proteoliposomes had an average diameter of 140, 149 and 160 nm determined from dynamic light scattering (Viscotek 802).

**AqpZ water permeability assay.** Liposome shrinkage was induced by rapidly mixing 1:1 (v/v) proteoliposomes/liposomes with osmolyte buffer (20 mM HEPES pH 7.5, 570 mM sucrose; a 285 milliosmolarity gradient) using a stopped-flow apparatus (Applied Photophysics SX20). Water transport/permeability was monitored by measuring the light scattering ( $\lambda = 600$  nm) of the preparation upon mixing at 8 °C. An increase in signal reflects liposome shrinkage. Light scattering ( $I$ ) data from repeated measurements ( $n = 5$ ) were fit to a single exponential rise equation to obtain the rate of water transport ( $k_{\text{wat}}$ ) using OriginPro 8.5.1.

$$L(t) = Ae^{-tk_{\text{wat}}} + L_0$$

where  $t$  is time in msec,  $A$  is a constant,  $k_{\text{wat}}$  is rate of water transport, and  $L_0$  is the maximum light scattering (typically one for normalized data).

**AmtB crystallization and structure determination.** AmtB was purified as described for native mass spectrometry experiments with one minor modification. AmtB was detergent exchanged into buffer H containing 0.5% C<sub>8</sub>E<sub>4</sub> using a Superdex 200 10/300 gel filtration column (GE Healthcare). Peak fractions of AmtB from gel filtration were pooled and concentrated to ~15 mg ml<sup>-1</sup> using a 50 kDa MWCO concentrator (Millipore). A tenfold molar excess of phosphatidylglycerol was added to AmtB before crystallization. Crystals of AmtB were grown at 20 °C in hanging drop plates with crystallization solution 15% PEG 4000, 0.8 M potassium formate, and 0.1 M sodium acetate pH 4.6. Hexagonal plate crystals appeared after one month.

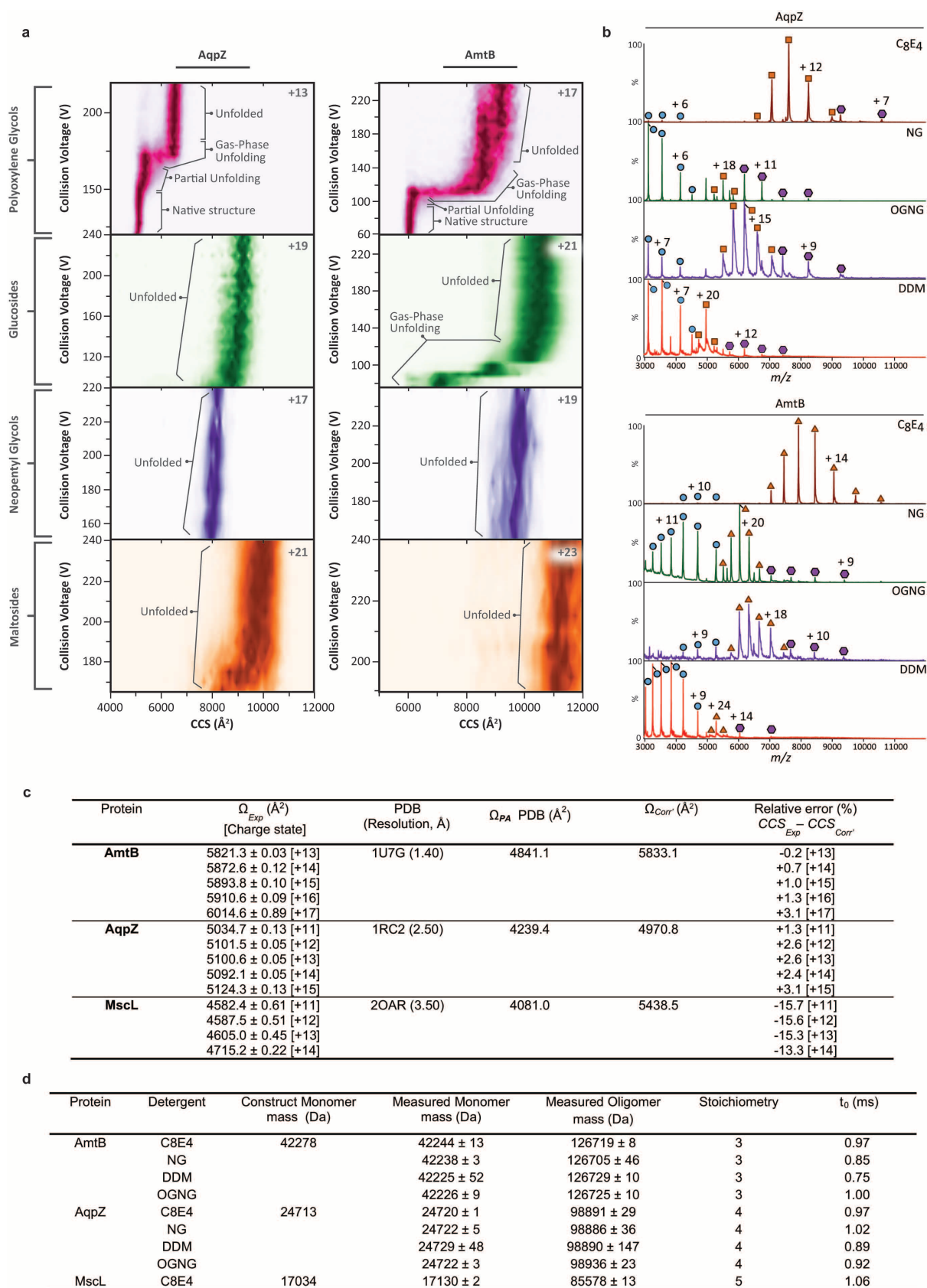
Single crystals were mounted with CrystalCap HT Cryoloops (Hampton Research, Aliso Viejo, CA) before flash frozen. Data were collected at Diamond Light Source beam line I04. Diffraction data collected at a wavelength of 0.97949 Å were processed with XDS<sup>65</sup> in apparent space group P6<sub>2</sub>2. Cell content analysis suggested two molecules in the asymmetric cell and attempts to obtain phases from molecular replacement using PHASER<sup>66</sup> with PDB: 1U7G resulted in placement of only one molecule with a missing layer of protein molecules. Twinning was suspected and the data were processed into lower symmetry space groups until molecular replacement could correctly place all protein molecules in the asymmetric cell, which turned out to be space group C22<sub>1</sub>. Data were submitted to the UCLA twin detection server<sup>67</sup> for twinning analysis that indicated partial twinning with  $\langle |L| \rangle = 0.463$  (untwinned = 0.5, perfectly twinned = 0.375) and  $\langle L_2 \rangle = 0.291$  (untwinned = 0.333, perfectly twinned = 0.200). Despite the presence of twinning, initial phases were found by molecular replacement using PHASER with PDB 1U7G placing six molecules in the asymmetric unit cell, and followed automated model building using PHENIX<sup>68</sup> and manual model building using COOT<sup>69</sup>. All model refinement was done with REFMAC<sup>70</sup> in twin mode. OMIT maps were generated first by deleting coordinates of the ligand before refinement. Refinement with twinning

applied resulted in a 10% drop in both  $R_{\text{work}}$  and  $R_{\text{free}}$ , consistent with twinning. Ramachandran plot of the refined structure gave 96% in favoured regions and 0.4% outliers. Figures were generated with PyMOL<sup>71</sup>.

- Gandhi, C. S., Walton, T. A. & Rees, D. C. OCAM: a new tool for studying the oligomeric diversity of MscL channels. *Protein Sci.* **20**, 313–326 (2011).
- Hilf, R. J. & Dutzler, R. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature* **452**, 375–379 (2008).
- Lee, D. J. *et al.* Gene doctoring: a method for recombineering in laboratory and pathogenic *Escherichia coli* strains. *BMC Microbiol.* **9**, 252 (2009).
- Drew, D., Lerch, M., Kunji, E., Slotboom, D. J. & de Gier, J. W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nature Methods* **3**, 303–313 (2006).
- Newby, Z. E. *et al.* A general protocol for the crystallization of membrane proteins for X-ray structural investigation. *Nature Protocols* **4**, 619–637 (2009).
- Laganowsky, A. *et al.* Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Sci.* **19**, 1031–1043 (2010).
- Laganowsky, A. *et al.* Atomic view of a toxic amyloid small oligomer. *Science* **335**, 1228–1231 (2012).
- Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook* 571–607 (Humana Press, 2005).
- Savage, D. F., O'Connell, J. D. III, Miercke, L. J., Finer-Moore, J. & Stroud, R. M. Structural context shapes the aquaporin selectivity filter. *Proc. Natl Acad. Sci. USA* **107**, 17164–17169 (2010).
- Sobott, F., Hernandez, H., McCommon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**, 1402–1407 (2002).
- Bush, M. F. *et al.* Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal. Chem.* **82**, 9557–9565 (2010).
- Mason, E. A. & Schamp, H. W. Mobility of gaseous ions in weak electric fields. *Ann. Phys.* **4**, 233–270 (1958).
- Shvartsburg, A. A. & Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **261**, 86–91 (1996).
- Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. Structural information from ion mobility measurements: effects of the long-range potential. *J. Phys. Chem.* **100**, 16082–16086 (1996).
- Benesch, J. L. P. & Ruotolo, B. T. Mass spectrometry: come of age for structural and dynamical biology. *Curr. Opin. Struct. Biol.* **21**, 641–649 (2011).
- Hall, Z., Politis, A. & Robinson, C. V. Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry. *Structure* **20**, 1596–1609 (2012).
- Ruotolo, B. T., Benesch, J. L., Sandercock, A. M., Hyung, S. J. & Robinson, C. V. Ion mobility-mass spectrometry analysis of large protein complexes. *Nature Protocols* **3**, 1139–1152 (2008).
- Stengel, F. *et al.* Dissecting heterogeneous molecular chaperone complexes using a mass spectrum deconvolution approach. *Chem. Biol.* **19**, 599–607 (2012).
- Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
- Jones, E. *et al.* Scipy: open source scientific tools for Python <http://www.scipy.org> (2001).
- Izzo, D., Ruciński, M. & Biscani, F. The generalized island model. *Studies in Computational Intelligence* **415**, 151–169 (2012).
- Izzo, D. PyGMO and PyKEP: open source tools for massively parallel optimization in astrodynamics (the case of interplanetary trajectory optimization). *Proceed. Fifth International Conf. Astrodynam. Tools and Techniques, ICATT* (2012).
- Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).
- Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
- Ulmschneider, J. P. & Ulmschneider, M. B. United atom lipid parameters for combination with the optimized potentials for liquid simulations all-atom force fields. *J. Chem. Theor. Comput.* **5**, 1803–1813 (2009).
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chemical Phys.* **79**, 926–935 (1983).
- Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chemical Phys.* **98**, 10089–10092 (1993).
- Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
- Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chemical Phys.* **126**, 014101 (2007).
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chemical Phys.* **81**, 3684–3690 (1984).
- Folch, J., Lees, M. & Sloane Stanley, G. H. A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.* **226**, 497–509 (1957).

62. Fine, J. B. & Sprecher, H. Unidimensional thin-layer chromatography of phospholipids on boric acid-impregnated plates. *J. Lipid Res.* **23**, 660–663 (1982).
63. Tan, B. K. *et al.* Discovery of a cardiolipin synthase utilizing phosphatidylethanolamine and phosphatidylglycerol as substrates. *Proc. Natl Acad. Sci. USA* **109**, 16504–16509 (2012).
64. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).
65. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
66. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
67. Padilla, J. E. & Yeates, T. O. A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr. D* **59**, 1124–1130 (2003).
68. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
69. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
70. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
71. Schrodinger, LLC *The PyMOL Molecular Graphics System, Version 1.3r1*, (2010).

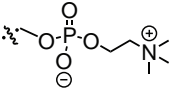
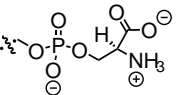
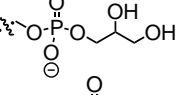
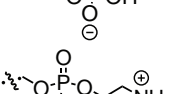
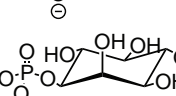
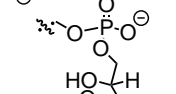
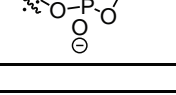




**Extended Data Figure 1 | Maintaining intact native membrane protein complexes in the mass spectrometer.** **a, b**, Gas-phase unfolding plots (left, 5 V steps) and mass spectra (right) of detergent stripped AqpZ and AmtB ions (charge inset) from different non-ionic detergent solutions; C<sub>8</sub>E<sub>4</sub> (pink), NG (green), OGNG (purple) and DDM (orange). Membrane protein complexes from NG, OGNG and DDM possessed CCS values substantially larger than

those calculated from crystal structures. **c**, Complete removal of C<sub>8</sub>E<sub>4</sub> at low-collision voltages reveals CCS values consistent with those calculated from their respective crystal structures. **d**, Reported are measured masses, standard deviations, and empirical  $t_0$  values used for direct CCS calculation of membrane proteins studied.

**a**

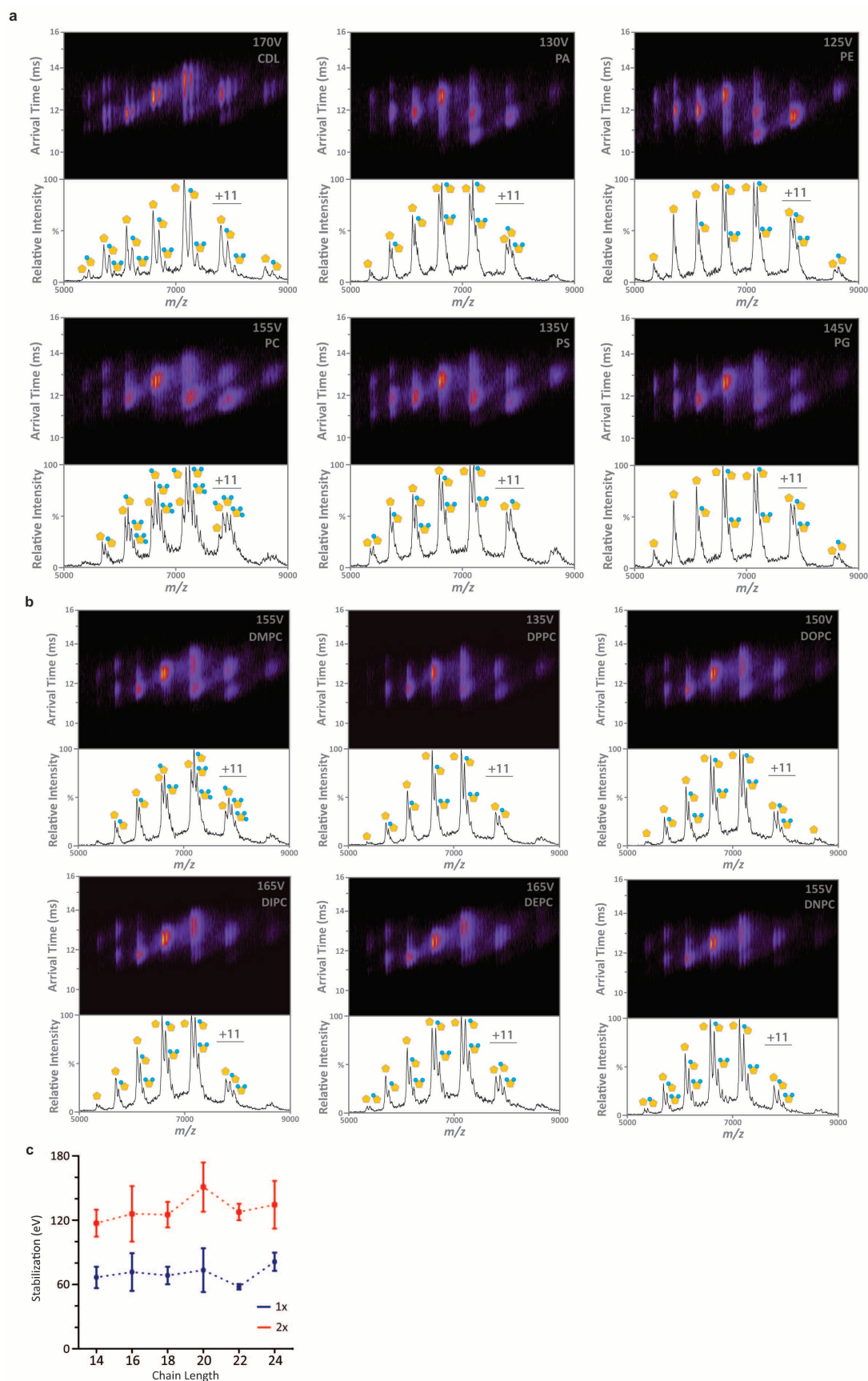
Phospholipid Name	Abbreviation	Average Molecular Weight	Phospholipid Head Group
1,2-dinervonoyl- <i>sn</i> -glycero-3-phosphocholine (24:1 Cis)	DNPC	954.4	
1,2-dierucoyl- <i>sn</i> -glycero-3-phosphocholine (22:1 Cis)	DEPC	898.3	
1,2-dieicosenoyl- <i>sn</i> -glycero-3-phosphocholine (20:1 Cis)	DIPC	842.2	
1,2-dioleoyl- <i>sn</i> -glycero-3-phosphocholine (18:1 Cis)	DOPC	786.1	
1,2-dipalmitoleyl- <i>sn</i> -glycero-3-phosphocholine (16:1 Cis)	DPPC	730.0	
1,2-dimyristoleyl- <i>sn</i> -glycero-3-phosphocholine (14:1 Cis)	DMPC	673.9	
1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-phosphocholine (16:0-18:1)	PC	760.1	
1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-phospho-L-serine (16:0-18:1)	PS	784.0	
L-α-phosphatidylglycerol ( <i>E. coli</i> )	PG	761.1	
1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-phosphate (16:0-18:1)	PA	696.9	
L-α-phosphatidylethanolamine ( <i>E. coli</i> )	PE	719.3	
1,2-dioleoyl- <i>sn</i> -glycero-3-phospho-(1'-myo-inositol-3'-phosphate) (18:1)	PI	977.1	
Cardiolipin ( <i>E. coli</i> )	CDL	1430.0	

**b**

Protein	Phospholipid	Protein Oligomer (μM)	Phospholipid (μM)	Detergent (μM)	Ratio (P:L:D)	# Bound	Lipid mass addition (Da)
AqpZ	CDL	7	70	16300	1:10:2300	1	1378.5 ± 12.5
	PE	7	348	-	1:50:2300	1	718.2 ± 18.5
	PG	9	87	-	1:10:1757	3	767.6 ± 11.1
	PS	9	383	-	1:40:1757	3	785.5 ± 25.6
	PC	7	132	-	1:18:2300	3	762.3 ± 14.0
AmtB	PA	7	449	-	1:61:2300	2	658.2 ± 8.0
	CDL	3	140	16300	1:45:5197	1	1417.9 ± 68.1
	PE	2	70	-	1:44:10395	2	688.7 ± 29.1
	PG	4	99	-	1:28:4559	4	756.9 ± 20.8
	PS	4	92	-	1:26:4559	3	759.9 ± 22.9
AmtB <sup>N72A/N79A</sup> MscL	PC	4	95	-	1:27:4559	3	756.5 ± 20.9
	PA	2	69	-	1:44:10395	3	664.5 ± 14.0
	PG	6	131	-	1:21:2558	3	763.1 ± 21.3
	CDL	17	175	8157	1:10:480	1	1430.0 ± 41.7
	PE	17	70	-	1:4:480	1	705.5 ± 25.8
	PG	17	131	-	1:8:480	2	778.4 ± 39.4
	PS	17	128	-	1:7:480	2	794.8 ± 22.0
	PC	17	132	-	1:7:480	3	776.4 ± 29.7
	PA	17	143	-	1:8:480	1	672.7 ± 36.5
	PI	17	52	-	1:6:480	4	957.3 ± 13.8
	DNPC	24	79	-	1:3:350	2	967.5 ± 29.7
	DEPC	24	56	-	1:2:350	2	913.6 ± 13.9
	DIPC	24	89	-	1:4:350	2	830.2 ± 52.9
	DOPC	24	95	-	1:4:350	2	793.7 ± 22.2
	DPPC	24	137	-	1:6:350	2	752.9 ± 52.2
	DMPC	24	74	-	1:3:350	2	695.6 ± 54.9

**Extended Data Figure 2 | Phospholipid abbreviations and optimization of phospholipid binding experiments.** **a**, Phospholipid abbreviations and headgroup structures. **b**, Protein to phospholipid to detergent ratios (P:L:D) for each membrane protein, number of resolvable phospholipids bound with their

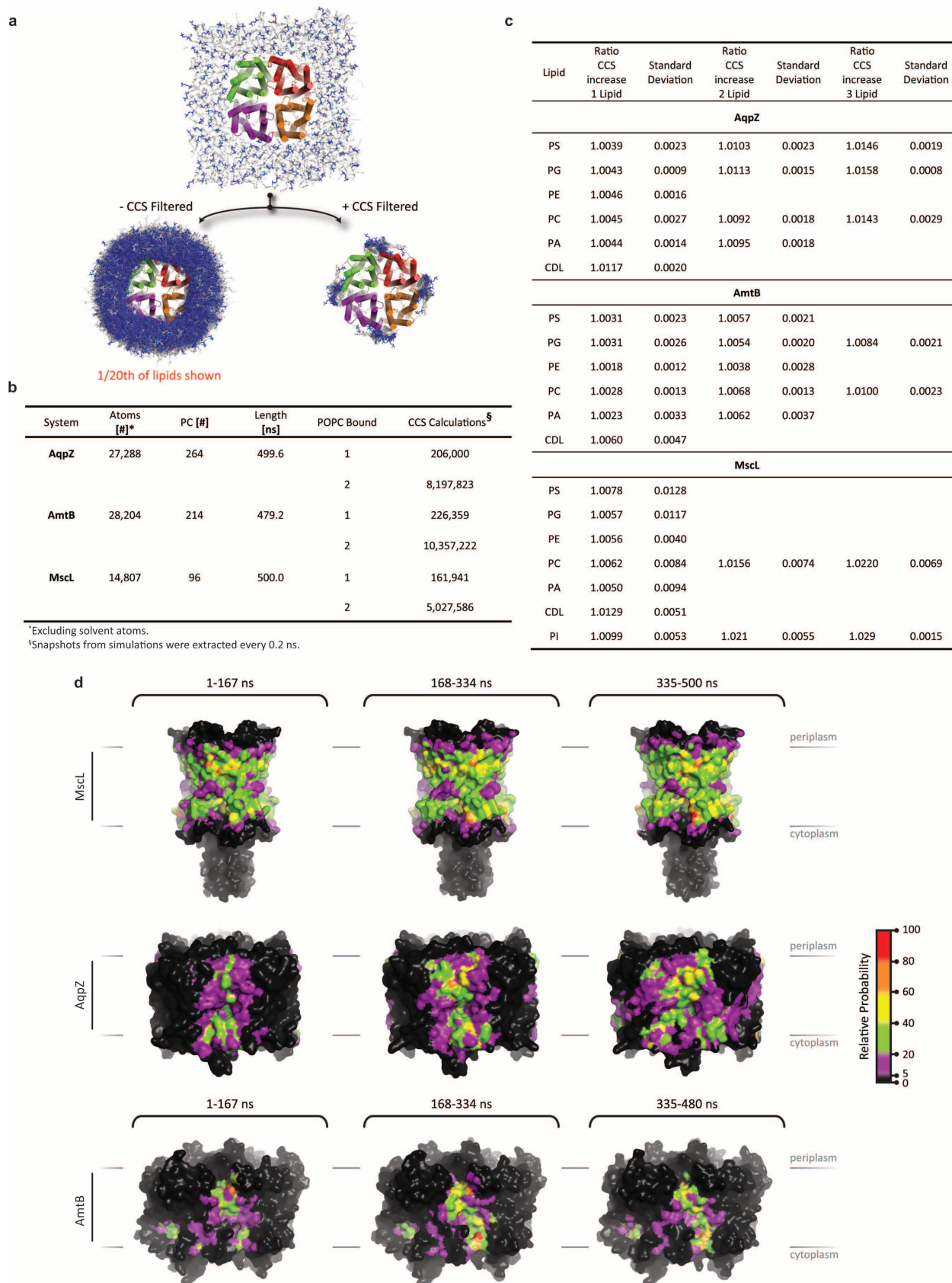
masses. Conditions were optimized empirically to maintain nanospray, sufficient mass spectral quality and phospholipid binding. Masses for one lipid bound to the protein complex were measured using MassLynx V4.1 software (Waters).



**Extended Data Figure 3 | Phospholipid binding to MscL reveals an insignificant impact on protein gas-phase stabilization independent of lipid alkyl chain length.** **a**, Representative mass and ion mobility spectra of MscL bound to various phospholipids species (inset top right). **b**, Representative mass

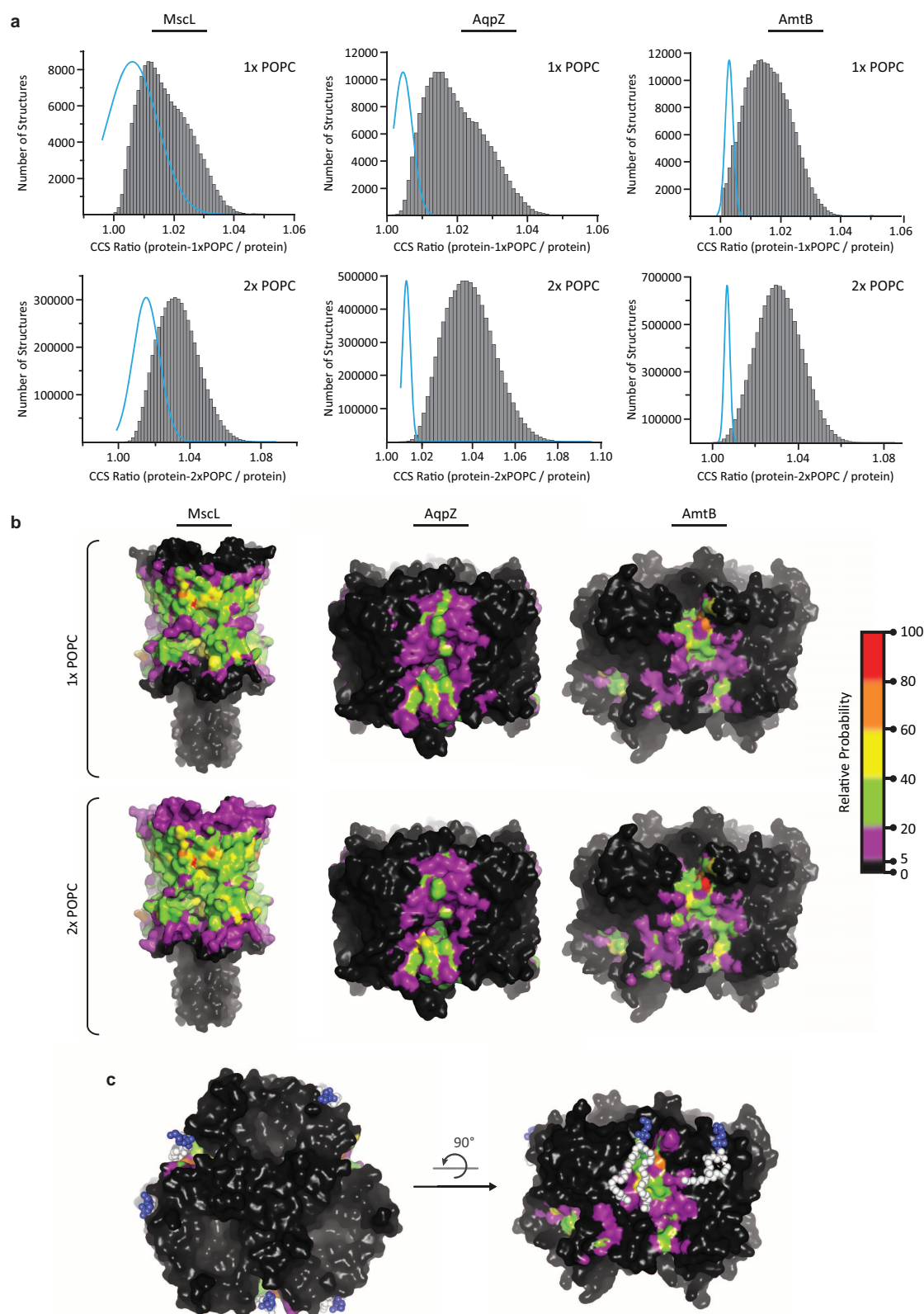
and ion mobility spectra of MscL bound to phospholipids of varying alkyl chain length. **c**, Cumulative stabilization of MscL is seen for one (blue) to two (red) bound lipid molecules. Reported are the means  $\pm$  s.e.m. from repeated measurements ( $n = 3$ ).





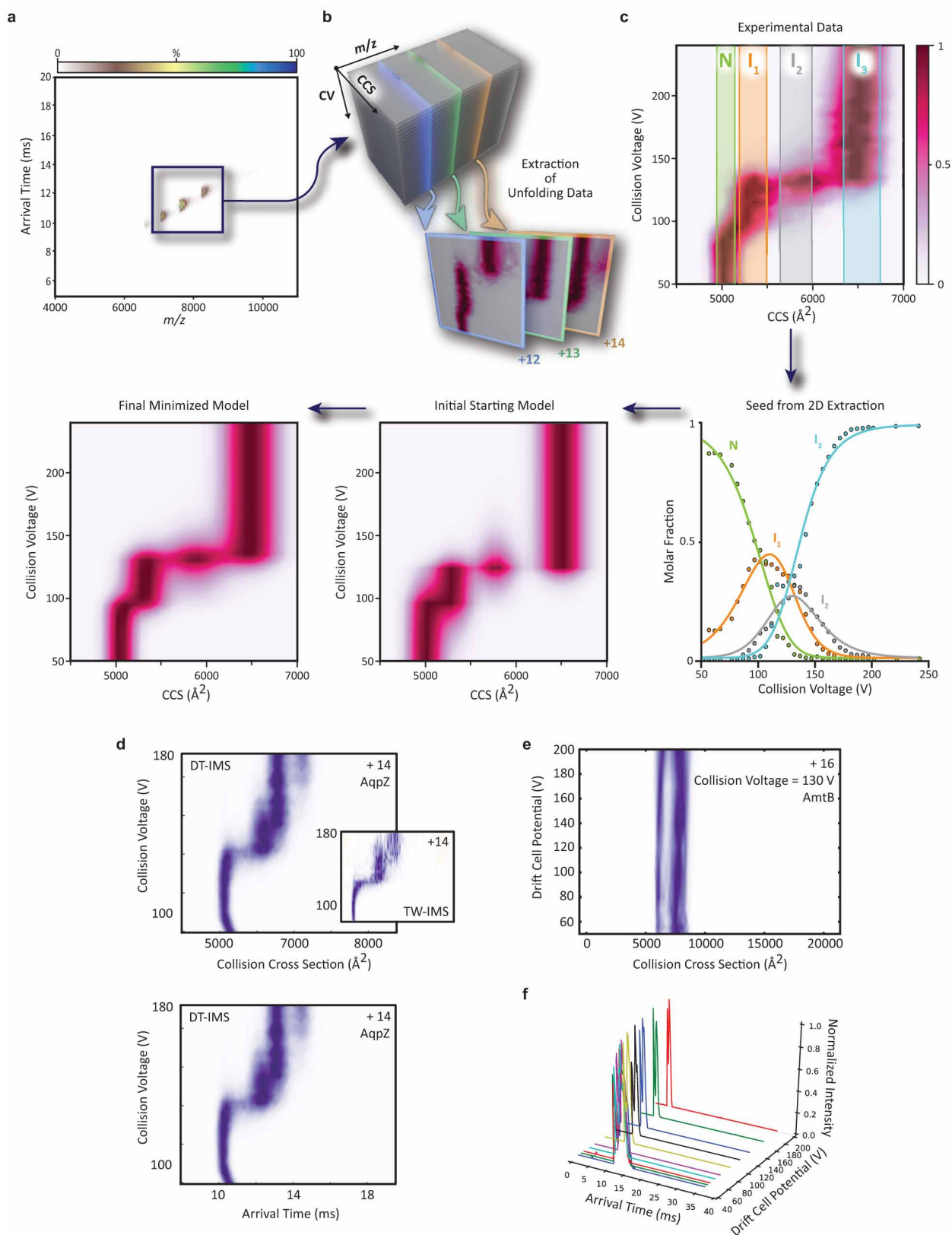
**Extended Data Figure 4 | Summary of statistics for molecular dynamics (MD) simulations followed by CCS filtering.** **a**, Representative filtering procedure of AqpZ in a PC bilayer from the MD (top). For each frame, lipids are extracted within 6 Å of the protein (bottom left) then filtered by the experimental CCS of AqpZ bound to a single PC molecule (bottom right);

see Methods. **b**, Summary of statistics. **c**, Ratios of phospholipid bound to apo CCS values for membrane protein complexes. Reported are the mean and standard deviation ( $n = 3$ ). **d**, MD filtered by CCS resulted in similar patches of PC molecules across different time points within the simulation indicating the systems were equilibrated.



**Extended Data Figure 5 | MD simulations filtered by CCS reveals probable one (1×) and two (2×) phospholipid binding sites.** **a**, For each candidate structure, protein and 1× or 2× PC molecule(s), the ratio of their calculated CCS values was determined (CCS ratio). This procedure generated a large number of candidate structures (grey bars) that were filtered using our CCS measurements (cyan line). The structures in grey that intercept this curve are essentially the ones selected as our most probable ensemble. **b**, The intersection between the simulated lipid complexes and the experimental data are then

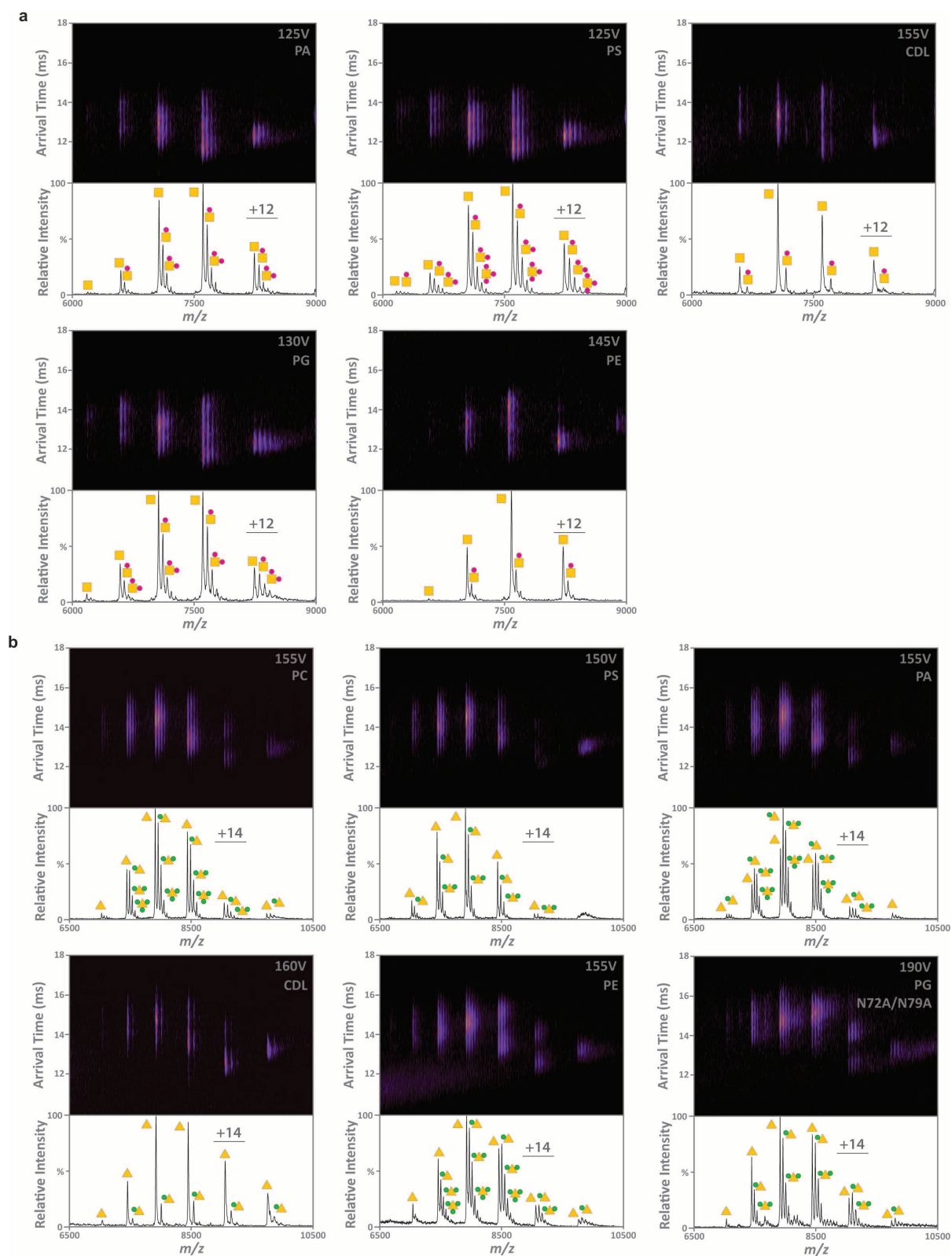
projected onto the surface of the protein to identify the most probable binding sites. Probable lipid locations for MscL resembled an annular belt, with no specific patches of lipids probably stemming from the relatively cylindrical geometry of this complex. By contrast, for AqpZ and AmtB the most probable location of the lipid molecules were localized to the interfacial regions between protein subunits, as well as other probable locations on individual monomers. **c**, X-ray derived PG (blue spheres with white tails) located at the subunit interfaces agrees with the predicted PC binding site.



**Extended Data Figure 6 | Modelling and quantification of gas-phase unfolding pathways.** **a**, Representative ion mobility mass spectra are collected over a range of collision voltages. **b**, Ion arrival time is converted to CCS before generating unfolding plots. **c**, Model fitting process from 2D to 3D data (see Methods). **d**, Both a Synapt2 modified with a linear drift cell (DT-IMS) and the commercially available travelling wave SynaptG2 (TW-IMS) produce qualitatively similar unfolding data. **e**, A contour plot representing the variance of CCS of two gas-phase unfolding species as a function of ion mobility drift cell

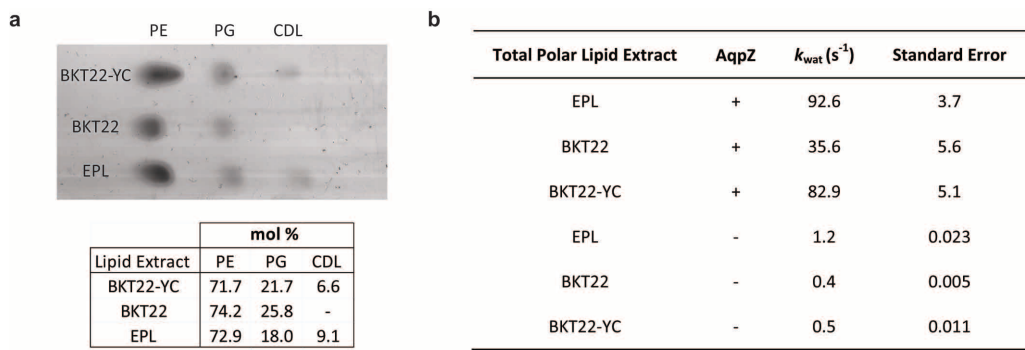
potential. **f**, Stacked plots of arrival time distributions for two gas-phase unfolding species as a function of drift cell potential. The lifetime of unfolding protein complexes in the drift tube ranges from 4 to 15 ms depending on the drift cell potential. No additional unfolding post activation occurs implying that the unfolding mechanism is not consistent with an irreversible unfolding model. Such a mechanism would predict time dependence on the population of unfolded species. By contrast the unfolding mechanism is well described by the reversible unfolding mechanism (see Methods).





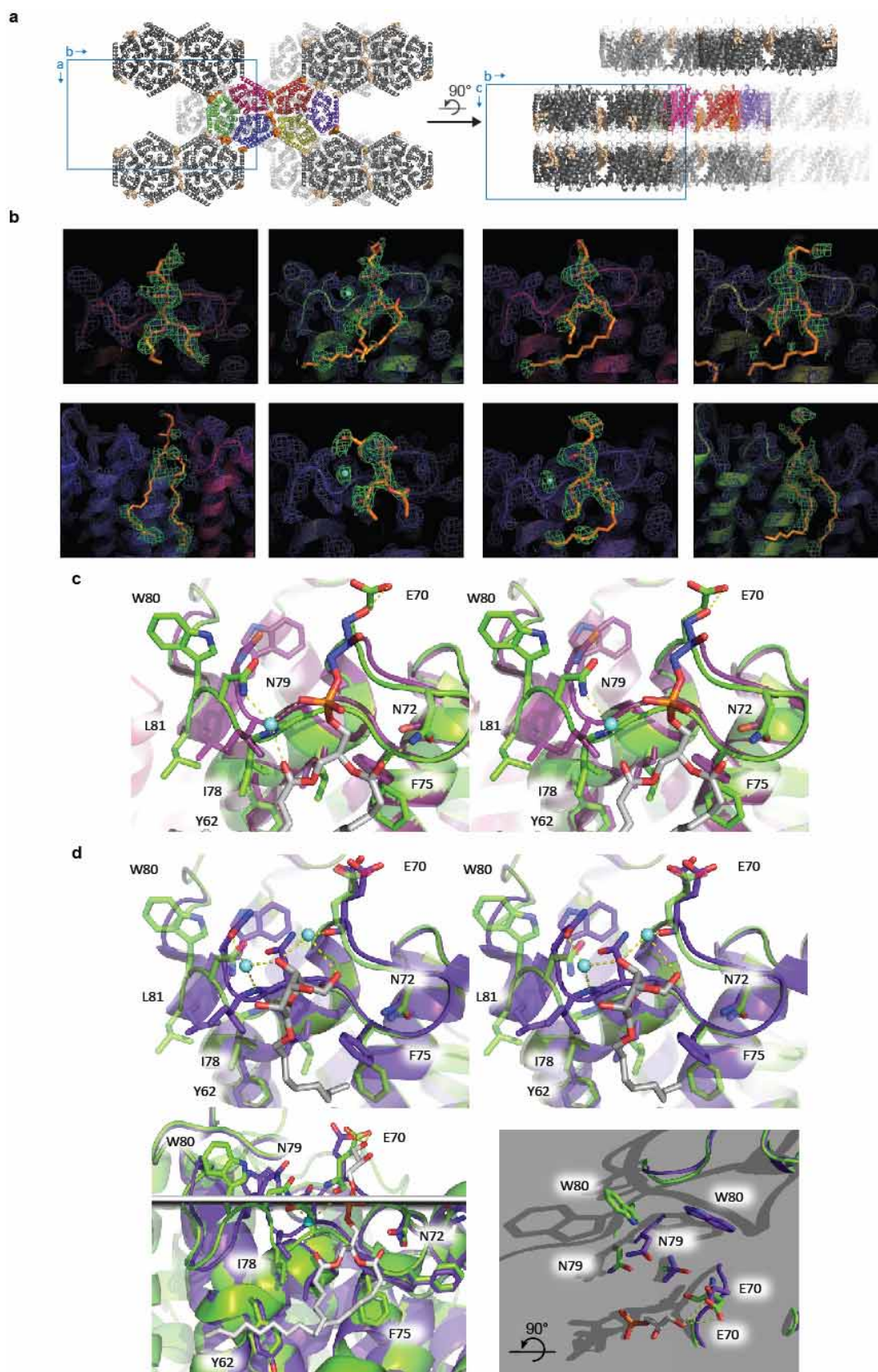
**Extended Data Figure 7 | AqpZ, AmtB and AmtB<sup>N72A/N79A</sup> bound to various phospholipid species. a,** Representative mass and ion mobility spectra

of AqpZ bound to phospholipids. **b,** Representative mass and ion mobility spectra of AmtB bound to phospholipids and AmtB<sup>N72A/N79A</sup> bound to PG.



**Extended Data Figure 8 | Summary of water permeability assays and analysis of lipid extracts.** **a**, HPTLC analysis of total polar lipid extract from wild-type *E. coli* (EPL), cardiolipin-deficient strain (BKT22), or BKT22 cells expressing ClsC and YmdB (BKT22-YC) to restore cardiolipin. Lipids were

quantified by densitometry. **b**, Reported are the rate constants ( $k_{\text{wat}}$ ) and standard error of replicates ( $n = 5$ ) for empty liposomes (–) and AqpZ proteoliposomes (+) reconstituted in differing *E. coli* lipid compositions.



#### Extended Data Figure 9 | Structural analysis of AmtB bound to PG.

**a**, Crystal packing with six AmtB (multicoloured) and eight PG (orange) molecules located in the asymmetric unit cell and symmetry related molecules shown in grey and light orange, respectively. **b**,  $F_o - F_c$  and  $2F_o - F_c$  electron density maps after refinement without lipid and near-lipid water molecule (if present) contoured at 2.0 and 1.0 sigma, respectively. **c**, Comparison of

AmtB bound to PG (green chain, this work) aligned with the AmtB structure (maroon chain, PDB: 1U7G). **d**, Structure overlay of AmtB-GlnK complex bound to octylglucoside (purple chain, PDB: 2NS1) aligned with AmtB bound to PG reveals a distinct conformational change. The lipid-water interface (grey plane) was determined from coordinates of phosphate atoms from bound PG.



**Extended Data Table 1 | Summary of X-ray data collection and refinement statistics**

	AmtB bound to PG
<b>Data collection</b>	
Space group	C222 <sub>1</sub>
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	116.2, 201.2, 232.5
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90
Resolution (Å)	40-2.3 (2.4-2.3) <sup>a</sup>
<i>R</i> <sub>merge</sub> (%)	12.4 (120.4)
<i>I</i> / $\sigma I$	13.7 (2.5)
Completeness (%)	99.9 (99.5)
Redundancy	11.0 (6.9)
<b>Refinement</b>	
Resolution (Å)	38.7-2.3 (2.36-2.3)
No. reflections	119,766
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	20.2 / 23.4 <sup>b</sup>
No. atoms	
Protein	15,856
Ligand/ion	293
Water	111
<i>B</i> -factors	
Protein	45.3
Ligand/ion	64.7
Water	29.5
R.m.s. deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.71

<sup>a</sup> Values in parentheses are for highest-resolution shell.<sup>b</sup> *R*<sub>free</sub> calculated using 5% of the data.

## CORRIGENDUM

doi:10.1038/nature13431

### Corrigendum: Realizing the promise of cancer predisposition genes

Nazneen Rahman

*Nature* **505**, 302–308 (2014); doi:10.1038/nature12981

In Figure 1 of this Article, the gene *RET* was incorrectly repeated on chromosome 12 and the genes *BRIP1* and *TRIM37* were omitted from chromosome 17. We thank Eric Song from Nanjing University, China, for bringing these inconsistencies to our attention. Figure 1 has been corrected online.

## ERRATUM

doi:10.1038/nature13463

### **Erratum: Recent advances in homogeneous nickel catalysis**

Sarah Z. Tasker, Eric A. Standley & Timothy F. Jamison

*Nature* **509**, 299–309 (2014); doi:10.1038/nature13274

In the first paragraph of this Review, the words ‘such as facile oxidative addition and ready access to multiple oxidation states’ were inadvertently repeated in the print version. The paper is correct online.



# CAREERS

**TURNING POINT** A move from China to America pays off for epigeneticist **p.179**

**@NATUREJOBS** Follow us on Twitter for the latest news and features [go.nature.com/e492gf](http://go.nature.com/e492gf)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



The band Bad Religion is fronted by evolution researcher Greg Graffin.

## SCIENCE AND THE ARTS

# Rock and research

*Scientists who moonlight as musicians get more from a gig than a fistful of cash.*

BY KAREN KAPLAN

Scientists are used to facing the music. Failure stares them down every time their grant is denied, their paper is rejected, their fellowship is declined or their experiment flops.

But many deal with the anguish of figuratively facing the music by literally making it. Although no formal statistics document the number of researchers who are professional or amateur musicians, anecdotal evidence suggests that they are legion. Those who cultivate a second life as musicians say that the misery of failed experiments, vetoed grant applications and rebuffed manuscripts falls away whenever they pick up their guitar, trumpet or drumsticks.

One might wonder why early-career researchers — who typically work 70–80-hour-plus weeks — spend what little spare time they have rehearsing, performing and recording

music when they could be, say, sleeping, or spending that time with their family. In part, say scientist-musicians, it is because performing live onstage or setting down tracks in the studio is akin to a powerfully addictive drug: it gives them such a rush that they must keep coming back for more. Indeed, research has confirmed that creating, performing and listening to music produces dopamine, the same neurochemical released during sex and other such pleasurable experiences (V. N. Salimpoor *et al.* *Nature Neurosci.* **14**, 257–262; 2011). But scientist-musicians also say that making music helps them to unwind and recalibrate. Rumination, recrimination or self-flagellation about what went wrong with the research programme or why the tenure application was turned down disappears when they focus on the next note or chord and the audience's response.

It has another advantage, too. For many who

pursue music, it feeds and bolsters their science. Freeing their mind from the tightly structured rigours of an experiment and the tedium of data collection to wander through fields of melodies and measures nurtures their scientific creativity. Days, even hours, after a particularly satisfying performance, writing stint or rehearsal session, a researcher often finds that she or he can suddenly devise a better way to approach a study, develop a stronger hypothesis for a manuscript or come up with a different rationale for a grant.

There is, of course, another, more fundamental reason for the time sacrifice: performing music onstage, writing and even rehearsing it is great fun.

“You get obsessed with it,” says Joseph LeDoux, a neuroscientist at New York University and singer-songwriter and guitarist for blues-rock band The Amygdaloids, who perform at clubs and other venues and have ▶

► recorded two original CDs and an extended play. “A lot of what I do for the band is write, and I’m a decent writer. It’s not the same kind of writing as a paper or a grant, but it’s freeing and releasing to write that way.”

The immediate audience feedback is a huge part of live music’s allure and a counterpoint to science’s slower pace, says Cliff Schweinfest, scientific review officer for the US National Cancer Institute in Rockville, Maryland, and trumpet player for the NIH Community Orchestra, a volunteer group of classical and Renaissance musicians whose day jobs are with the US National Institutes of Health (NIH) and other federal agencies. “Music gives you an instantaneous pay-off,” he says. “That’s very satisfying — it’s a good complement to science, which has a much slower pace.”

The music–science interplay can have real benefits. One study, for instance, has found that people who form companies and file patents are much more likely to be involved in the arts than are those who have never done these things (R. LaMore *et al. Econ. Dev. Q.* 27, 221–229; 2013). Long-term participation may enhance creative potential in science and technology, and scientists who are musicians are likely to be highly successful, publishing provocative papers or producing many patentable inventions, says psychologist Robert Root-Bernstein from Michigan State University in East Lansing, and an author of the study. “They are more successful by any set of criteria, including publishing and grants,” he says. “They are more innovative. A lifelong exposure to music helps to foster a more creative approach to work.”

## BLURRED BOUNDARIES

So tightly intertwined are science and music for evolutionary researcher Greg Graffin, who fronts the punk-rock outfit Bad Religion in Los Angeles, California, that the lines between them often blur. Lead singer and songwriter for the band since he was 15, Graffin perceives a profound connection between the anti-establishment, alienative nature of punk music and the demands of scientific research for rational thinking and for challenging existing doctrine. The mindset has piloted both his science and his music for the past 35 years. “That was the way I was thinking when I was 16 — to challenge authority to make headway in this world,” he says. “It’s worked for me in punk, and it’s what I do as a scientist.” Of the four books he has published, *Anarchy Evolution: Faith, Science, and Bad Religion in a World Without God* is perhaps the most iconoclastic, examining the conflict between religion and science.

Hugely successful as a punker — Bad Religion headlined iconic clubs such as Whiskey a Go-Go in West Hollywood, California, when he was 17 — Graffin slowed his research career for a while, fitting in a doctorate, a postdoc and stints at the bench between tours and records. But he could never bring himself to leave science completely for music, lucrative though his



Punk band the Descendents release albums around research and family commitments.

music career has been. He is now a lecturer in evolution at Cornell University in Ithaca, New York, and is preparing to publish a book on the biological basis of species coexistence. He teaches and conducts research in the autumn and spring, and reserves summers for touring, writing and recording — because even now, decades later, music still holds a seductive allure.

“To this day, there’s nothing more rewarding than playing concerts and just seeing the people mouth the words and sing along,” he says. “There’s such a dichotomy — the lights, the stage, the sensation of the moment — and the inverse of that, the quiet solitary work of data collection.”

Although his music and his science are closely knit in his mind, Graffin tries to keep them separate, especially for his students, in an effort to preserve his scientific identity. “In a lecture, I don’t even talk about music, but [students] come up after class,” he says. “I don’t want to promote that part of my life — in a sense they’re indulging me, to see me as a scientist. I’m walking the line between the two worlds.”

Plant geneticist Milo Aukerman also straddles an oft-indistinct boundary between research and music. Long-time frontman and lead singer for Los Angeles-based early punkers the Descendents, Aukerman followed a path similar to Graffin’s, sandwiching his graduate programme and postdoc between tours and recording sessions. After completing his postdoc in 1999 — it took six years, thanks to a series of leaves for touring and recording — Aukerman joined DuPont in Wilmington, Delaware, in 2001. Then music came knocking again. “I hadn’t been at DuPont for more than a year when the band called. I said, ‘I can’t

tour,’ and they said, ‘How about a record?’, and I said ‘I could squeeze that in,’” he recalls. The CD was released in 2004 to significant acclaim, but Aukerman had decided it was time to focus again on his science — until 2010, when the muse called him back once more. Now, he joins the band for weekend shows and brief international tours that he fits in during holidays.

## COMPLEMENTARY CURES

Music and science, he says, are different sides of the same coin — they both draw on creative energies and abilities. “From the very beginning of my interest in biology, it was about what you could do. This was a raw substance, this was like something you sculpted, that you could create with,” Aukerman says. “It’s the same way in music. Having innovated — writing a song, interpreting a riff in a different way — brings the same exhilaration that you get from science.”

And the research can inform the music. The Amygdaloids’ original tunes are about the brain or the mind in some way — and LeDoux finds that the music affects his work in a positive fashion. “I’ve got much better at being a performer — I’m more relaxed now, more spontaneous, and that also helps flipping back to science — I’m more willing to free-associate,” he says, adding that being a rocker has improved his performance on the podium when he gives talks. “During a science lecture, you might be in your own mind, facing the screen, looking for the next slide, but with music, you’ve got to be right there — onstage as a musician, it’s all about grabbing the audience. And I’ve become much more conscious of that in my work.”

For some, the music offers more direct insights: it is part and parcel of the research itself. Thomas Scelo, an acoustics consultant and classical guitarist, knows that the connection between his music and his research is substantive, even crucial. The director of Marshall Day Acoustics in Hong Kong, Scelo works with cities, private companies, architects and others

REBECCA REED